

Universidad Autónoma de Madrid

Departamento de Biología Molecular



**ESTUDIO DE LA ESPECIFICIDAD
FUNCIONAL DE AID EN EL PROCESO DE
HIPERMUTACIÓN SOMÁTICA**

Tesis Doctoral

Pablo Pérez Durán

Madrid, 2012



Memoria presentada por Pablo Pérez Durán, Licenciado en Biología, para optar al grado de Doctor en Ciencias por la Universidad Autónoma de Madrid.

Esta Tesis Doctoral ha sido realizada en el laboratorio de Hipermutación de ADN y Cáncer del Centro Nacional de Investigaciones Oncológicas (CNIO) bajo la dirección de la Doctora Almudena R. Ramiro.

Para su realización se contó con financiación del Ministerio de Ciencia e Innovación (MICINN) a través del subprograma de Formación de Personal Investigador (FPI) y el European Research Council (ERC).

Índice

Abstract	3
Introducción	7
Inmunidad innata e inmunidad adaptativa.....	9
Diversificación de anticuerpos	9
Diversificación primaria	9
Diversificación secundaria	10
La desaminasa inducida por activación (AID).....	11
Procesamiento del uracilo por UNG	13
Reconocimiento del apareamiento U:G por Msh2-Msh6	14
AID y la reacción de cambio de isotipo	16
Efectos adversos de la actividad de AID en células B.....	16
Mecanismos reguladores de la actividad de AID.....	17
Expresión y actividad de AID en células no B	20
Objetivos.....	23
Resultados.....	27
Desarrollo de un ensayo reportero para monitorizar las mutaciones introducidas por AID.....	29
Generación de un casete reportero para monitorizar la actividad de AID	29
Validación del casete reportero en un sistema inducible.....	31
Identificación de <i>loci</i> permisivos a la actividad de AID	35
Identificación de <i>loci</i> transcripcionalmente activos.....	35
Análisis de los <i>loci</i> transcripcionalmente activos	37
Determinación del sitio de inserción	39
Análisis de la contribución de UNG a la especificidad de secuencia de la SHM	44
La inhibición de UNG bloquea CSR y la aparición de revertientes.....	45
La inhibición de UNG incrementa la frecuencia de mutaciones	47
Análisis de las mutaciones inducidas por AID mediante secuenciación masiva	48
La sobreexpresión de UNG reduce el CSR y aumenta la aparición de revertientes.....	53
La sobreexpresión de UNG modifica el patrón de mutaciones inducidas por AID	54
UNG regula el balance entre reparación y mutación en <i>hotspots</i> de AID	56

Discusión	59
Ensayo de revertientes y NGS para el análisis de la actividad de AID.....	61
<i>Locí</i> susceptibles a la actividad de AID.....	63
Especificidad local de secuencia de AID.....	66
Contribución de UNG a la especificidad local de AID.....	68
Conclusiones	73
Materiales y métodos	77
Cultivos celulares.....	79
Construcción de vectores.....	79
Transfección estable de células NIH-3T3.....	80
Transducción retroviral de células de ratón.....	81
Análisis de la expresión de ARN mensajero por PCR cuantitativa a tiempo real (RT-qPCR)	82
PCR inversa	82
Amplificación por PCR de orangeSTOP y secuenciación por <i>Sanger</i>	83
Identificación de mutaciones en orangeSTOP por secuenciación masiva.....	84
Amplificación de S μ e identificación de mutaciones por secuenciación masiva	85
Southern blot	85
Ensayos de ganancia y pérdida de función de UNG en células B.....	86
Bibliografía	87

Abstract

An effective immune response relies on the secondary diversification of antibodies, and is achieved by two different processes, namely somatic hypermutation (SHM) and class switch recombination (CSR). Activation induced deaminase (AID) triggers both processes by deaminating cytosines on the DNA of immunoglobulin genes. Cytosine deamination results in the generation of a U:G mismatch that can be processed by different pathways giving rise to a double strand break in the case of CSR or a nucleotide substitution in SHM. Uracil-N-Glycosylase (UNG), an enzyme normally involved in base excision repair, is required both for CSR and SHM. The mechanisms that determine AID specificity are not well understood and in fact this enzyme can act on genes other than immunoglobulin genes, generating mutations and chromosomal translocations with oncogenic potential.

To study AID locus specificity genome-wide we screened for susceptible genes making use of a promoterless construct which contains a dual fluorescence reporter cassette. Although random integration was expected, we observed a striking bias towards integration into a sequence of retroviral origin. Moreover, we found that all *loci* susceptible to AID activity integrated the reporter construct in this same site.

To assess the role of UNG in AID local sequence specificity we performed loss- and gain-of-function experiments and next generation sequence analysis of AID-induced mutations. We have shown that AID has a more stringent specificity than previously recognized and that UNG activity can promote both error-free and error-prone repair pathways to process AID-induced U:G mismatches. Moreover, the balance between these two pathways is partially determined by primary DNA sequence, with individual hotspots showing differential preference for one of them. Together, this work reveals UNG as a new molecular layer that shapes the specificity of AID in a sequence dependent fashion. Our results have important implications in the understanding of antibody secondary diversification and the contribution of AID to cancer development.

Introducción

Inmunidad innata e inmunidad adaptativa

El sistema inmune surge en la naturaleza como un sistema de protección de los organismos frente a agentes patógenos. En animales vertebrados este sistema es enormemente sofisticado y comprende dos tipos de respuesta inmune: innata y adaptativa. La inmunidad innata constituye una primera barrera de defensa rápida y eficiente, pero poco específica. La inmunidad adaptativa proporciona una barrera más especializada frente a los patógenos y además permite el desarrollo de la memoria inmunológica, que hace posible la eliminación del patógeno en futuras infecciones de manera más rápida y eficaz. Las células efectoras de la inmunidad adaptativa son los linfocitos o células T y B. Su actividad se basa en la expresión de receptores para el antígeno, llamados TCR (del inglés *T Cell Receptor*) en células T y anticuerpos o inmunoglobulinas en células B, que son los encargados de reconocer estructuras moleculares concretas (antígeno) presentes en el patógeno. La variabilidad de inmunoglobulinas es virtualmente ilimitada, lo que permite reconocer y combatir cualquier infección a la que el organismo pueda verse expuesto.

Diversificación de anticuerpos

Diversificación primaria

La molécula de inmunoglobulina, o anticuerpo en su forma secretada, se compone de dos pares de cadenas idénticas, llamadas cadena ligera (IgL, del inglés Immunoglobulin Light chain) y cadena pesada (IgH, del inglés Immunoglobulin Heavy chain). El extremo N-terminal de IgH e IgL configuran la región variable del anticuerpo responsable de reconocimiento antigénico, mientras que el extremo C-terminal de IgH constituye la región constante, que determina la vía celular de eliminación del antígeno (Fig. 1).

La diferenciación de las células B ocurre en la médula ósea y permite la generación de un repertorio primario de anticuerpos de gran variabilidad. La capacidad de codificación de información presente en el genoma no puede explicar esta gran diversidad sino que se consigue gracias a un mecanismo independiente de antígeno denominado recombinación V(D)J, homólogo al que diversifica el TCR en células T (Schatz, 2004). Los genes de

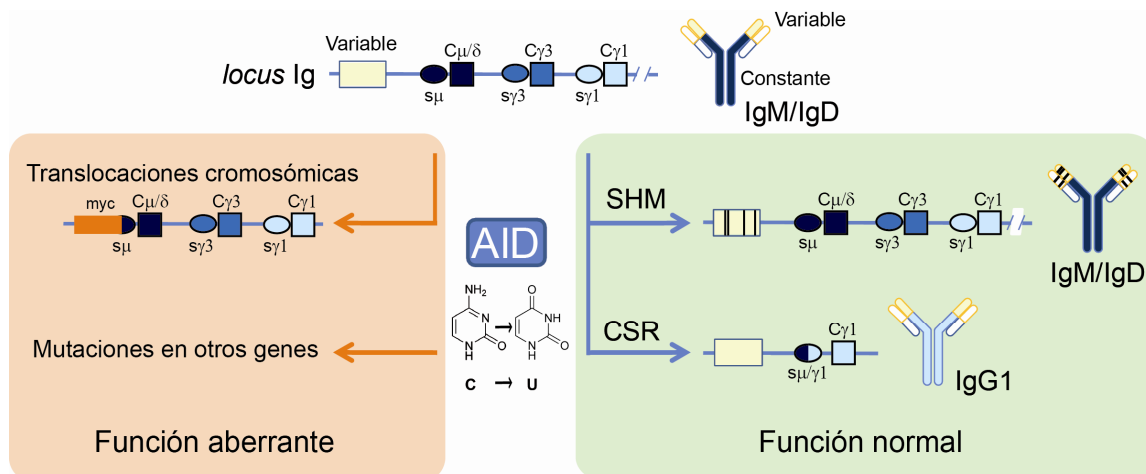


Figura 1: Actividad de AID en la reacción de centro germinal

En células B maduras el gen que codifica para la cadena pesada de inmunoglobulinas IgH (esquema superior) está constituido por una región en 5' que ha sufrido un reordenamiento V(D)J productivo (cuadrado amarillo) codificante para la región variable de reconocimiento antigénico de la molécula de anticuerpo y por una región en 3' que codifica para un número variable de regiones constantes (cuadrados azules) precedidas por regiones altamente repetitivas no codificantes, o regiones de *switch* (óvalos azules). El gen IgH puede sufrir un proceso de diversificación secundaria en los centros germinales como consecuencia del reconocimiento antigénico. Los dos mecanismos moleculares que constituyen este proceso son la SHM y la CSR. En la SHM se introducen mutaciones puntuales (rayas negras) en la región variable de la inmunoglobulina, mientras que la CSR consiste en un proceso de recombinación entre dos regiones de *switch* gracias al cual el anticuerpo expresa una cadena constante con un isotipo diferente (en la figura se representa IgG1). Ambas reacciones se inician gracias a la desaminación de citosinas en la cadena variable o en las regiones de *switch* mediada por AID. Además de esta función en la diversificación de anticuerpos la actividad de AID también tiene efectos adversos como pueden ser las translocaciones cromosómicas entre c-myc e IgH o la acumulación de mutaciones en otros genes distintos de las inmunoglobulinas.

inmunoglobulinas están compuestos por una colección de segmentos génicos denominados V (variable), J (unión, del inglés *joining*) y en el caso de la cadena pesada D (diversidad). La recombinación V(D)J consiste en el reordenamiento somático de estos segmentos para dar lugar en cada célula B a una yuxtaposición VDJ única. Los segmentos que intervienen en el proceso se escogen al azar, por lo que la variabilidad combinatoria de las inmunoglobulinas generadas es enorme. Se ha estimado que la recombinación V(D)J puede generar 10^5 - 10^6 anticuerpos distintos con diferentes especificidades (Di Noia and Neuberger, 2007). Estos anticuerpos aún pueden sufrir un proceso adicional de diversificación secundaria en los centros germinales después del reconocimiento antigénico, como se explicará a continuación.

Diversificación secundaria

A diferencia de los receptores para el antígeno de células T, los genes de inmunoglobulinas pueden llevar a cabo un proceso de diversificación secundaria, que tiene lugar en los centros germinales, como consecuencia del reconocimiento antigénico por parte de las células B. Esta

reacción es totalmente necesaria para poder generar anticuerpos de alta afinidad y con funciones efectoras diversas. Los dos mecanismos moleculares responsables de la remodelación del repertorio de anticuerpos son la hipermutación somática (SHM, del inglés *Somatic HyperMutation*) y la recombinación de cambio de isotipo (CSR, del inglés *Class Switch Recombination*).

La SHM implica la introducción de mutaciones puntuales en el ADN que codifica para la región variable (VDJ) de las inmunoglobulinas (Fig. 1). Si estas mutaciones incrementan la afinidad por el antígeno la célula B que las ha adquirido sufrirá un proceso de expansión clonal y se diferenciará en células B de memoria y células plasmáticas (secretoras de anticuerpos). Si por el contrario las mutaciones son deletéreas o resultan en anticuerpos de menor afinidad la célula sufrirá apoptosis. Este proceso se denomina maduración por afinidad y tiene como resultado final un aumento en la afinidad del anticuerpo por el antígeno correspondiente (Di Noia and Neuberger, 2007; Rajewsky, 1996).

La CSR es una reacción de recombinación específica de región por la que las regiones constantes $C\mu$ y $C\delta$ (que codifican para las cadenas pesadas de los isotipos IgM e IgD) se reemplazan por las regiones $C\gamma$, $C\epsilon$ o $C\alpha$ (que codifican para los isotipos IgG, IgE e IgA, respectivamente) situadas en el extremo 3' del locus de inmunoglobulinas (Fig. 1). La recombinación se produce entre dos regiones de *switch* distintas, que son secuencias no codificantes y altamente repetitivas que preceden a cada región constante. De esta forma, una misma región variable N-terminal del anticuerpo se combina con cadenas constantes distintas, adquiriendo así funciones efectoras diferentes (Stavnezer, 1996; Stavnezer et al., 2008). Gracias a estos dos procesos, el repertorio de anticuerpos incrementa su afinidad y su diversidad de respuesta, siendo fundamental para una correcta respuesta inmune.

La desaminasa inducida por activación (AID)

La SHM y CSR son iniciadas por la misma enzima, la desaminasa inducida por activación (AID, del inglés *Activation-Induced cytidine Deaminase*), que produce la desaminación de citosinas en la región variable o en las regiones de *switch* del locus de inmunoglobulinas respectivamente. La eliminación de un grupo amino en una citosina genera un desapareamiento U:G en el ADN. El tipo de sustituciones nucleotídicas inducidas en SHM y la generación de roturas de doble cadena en el ADN (DSBs, del inglés *Double Strand Breaks*), que son el

sustrato de la CSR, depende de cómo sea reconocido y procesado este desapareamiento inicial.

AID fue descubierta en 1999 mediante un análisis de genes diferencialmente expresados en una línea de células B de ratón (CH12F3) al ser estimulada a hacer CSR (Muramatsu et al., 1999). Posteriormente se observó que AID era imprescindible tanto para SHM como para CSR en ratones y humanos (Muramatsu et al., 2000; Revy et al., 2000). Estructuralmente, AID es una proteína de 198 aminoácidos con cuatro dominios funcionales descritos. El dominio catalítico de desaminación se localiza en una región central de la proteína. El extremo N-terminal tiene una alta concentración de aminoácidos básicos y se ha propuesto como posible dominio de localización nuclear (Ito et al., 2004) y además, se ha descrito un dominio de dimerización cercano a esta región. En el extremo C-terminal existe un dominio NES (del inglés, *Nuclear Export Signal*) de exportación nuclear (Ito et al., 2004; McBride et al., 2004). AID tiene una gran homología con APOBEC1, una enzima responsable de editar el ARN mensajero (ARNm) de la apolipoproteína B, por lo que inicialmente se propuso que AID podría realizar su función en la diversificación de anticuerpos mediante la edición de moléculas de ARN que serían las encargadas de introducir las mutaciones en SHM o producir los DSBs necesarios para la CSR (Muramatsu et al., 2000). No obstante, esta hipótesis se ha abandonado en la actualidad ya que numerosas evidencias indican que AID inicia la SHM y la CSR mediante la desaminación de citosinas directamente en el ADN (Chaudhuri et al., 2003; Maul et al., 2011; Petersen-Mahrt et al., 2002; Pham et al., 2003; Ramiro et al., 2003). En particular, mediante estudios *in vitro* se demostró que AID es capaz de desaminar citosinas presentes en ADN de cadena sencilla (ADNcs) pero no en ARN, ADN de cadena doble (ADNcd) o en híbridos ADN-ARN (Bransteitter et al., 2003; Chaudhuri et al., 2003; Dickerson et al., 2003; Pham et al., 2003). En estos experimentos también se observó que AID desamina preferentemente citosinas que estén incluidas en “puntos calientes” mutacionales o *hotspots* definidos por el motivo degenerado WRC (W=A/T, R=A/G) (Bransteitter et al., 2004; Pham et al., 2003).

Como se ha indicado anteriormente, la SHM implica la introducción de mutaciones en los segmentos V(D)J reordenados y se inicia gracias a la actividad de AID. El modelo de desaminación en el ADN, propuesto por el laboratorio de M.S. Neuberger, explica la aparición de todas las mutaciones detectadas en SHM. La desaminación de una citosina en el ADN genera un desapareamiento U:G. Si la célula replica su ADN antes de haber corregido este desapareamiento, la maquinaria de replicación introducirá una adenina enfrente del uracilo. Por lo tanto, el par original C:G se convertirá en un par T:A en una de las copias del ADN, generándose así transiciones del tipo C→T -o G→A si la desaminación ocurre en una C de la

hebra no transcrita- (Fig. 2). Alternativamente, si el desapareamiento U:G es reconocido por proteínas implicadas en la reparación por escisión de base (BER, del inglés *Base Excision Repair*) o en la reparación de desapareamientos (MMR, del inglés *MisMatch Repair*), se generarán el resto de posibles mutaciones (transversiones en pares G:C y mutaciones en pares A:T). La proteína encargada de reconocer esta lesión en la vía de BER es la glicosilasa de uracilos UNG, mientras que en la vía de MMR es el dímero Msh2-Msh6. Numerosas evidencias, detalladas en los siguientes capítulos, han validado este modelo.

Procesamiento del uracilo por UNG

Los uracilos en el ADN son reconocidos como una base extraña y normalmente son reparados mediante BER. Existen cuatro glicosilasas de ADN descritas en animales vertebrados que son capaces de escindir uracilos: UNG, SMUG1, TDG y MBD4 (Visnes et al., 2009). De todas ellas UNG es la única que tiene una función esencial en la diversificación de anticuerpos (Di Noia et al., 2006; Visnes et al., 2009)(di noia EMBO 06, Visnes-Krokan Philos Trans B 09). La importancia de UNG en este proceso se ha demostrado usando distintos sistemas experimentales. En primer lugar se transfectaron células B de pollo (DT40) con un inhibidor específico de UNG (Ugi) producido por el fago PSB2 y se analizaron los patrones de SHM. En células Ugi⁺ la proporción de transiciones en pares G:C en la región variable λ del gen de inmunoglobulinas estaba drásticamente aumentada con respecto a las células control y la proporción de transversiones estaba disminuida (Di Noia and Neuberger, 2002). Por otra parte también se observó un cambio en el patrón de mutaciones en ratones deficientes para UNG. En este caso se analizó la región variable en células B de centro germinal de placas de peyer (tejido linfoide asociado a la mucosa intestinal donde se produce una activación constitutiva de las células B) y se observó que la mayoría de las mutaciones detectadas eran transiciones en pares G:C (Rada et al., 2002). Finalmente, en humanos con mutaciones inactivantes de UNG también se observó una modificación similar en el patrón de mutaciones en la región variable de IgM (Imai et al., 2003). Estos resultados indicaban que UNG es necesaria para el proceso de SHM, en particular para la generación de transversiones en pares G:C. Según el modelo expuesto en la figura 2, el incremento en el porcentaje de transiciones en pares G:C observado en estos sistemas se debería a que en ausencia de la actividad de UNG el desapareamiento U:G sólo puede ser reconocido por la maquinaria de replicación o por el dímero Msh2-Msh6, que como veremos más adelante es el principal responsable de la introducción de mutaciones en pares A:T.

En la ruta de BER canónica, el sitio abásico generado por UNG se repara gracias a la acción de la ADN polimerasa β (Pol β). Se ha propuesto que las polimerasas TLS (del inglés *TransLesion Synthesis*), que están implicadas en la síntesis de ADN frente a lesiones, podrían participar en SHM sustituyendo a Pol β en BER (Weill and Reynaud, 2008). Las polimerasas TLS no son tan precisas en la incorporación de nucleótidos como Pol β . Este es el caso de Rev1, una polimerasa muy eficiente en la incorporación de nucleótidos, preferentemente desoxicitidinas, frente a un sitio abásico (Prakash et al., 2005), lo que daría lugar a una transversión. Se observó que en células B de ratón y de pollo deficientes para Rev1 este tipo de mutaciones estaba reducido en un 50% en la región variable de la inmunoglobulina (Jansen et al., 2006; Ross and Sale, 2006). Debido a que la reducción es parcial, otras polimerasas pueden estar implicadas en la generación de transversiones (Weill and Reynaud, 2008). También se ha propuesto la participación de otras polimerasas con distinta preferencia de incorporación de nucleótidos en la reparación del sitio abásico generado por UNG, como por ejemplo la ADN polimerasa θ (Di Noia and Neuberger, 2007; Weill and Reynaud, 2008), que podría generar tanto transversiones como transiciones en pares G:C (Fig. 2).

No obstante, se desconocen los mecanismos por los que la actividad de UNG, cuya función canónica inicia la ruta de reparación mediada por BER, se pervierte dando lugar a la generación de mutaciones en el contexto de la diversificación de anticuerpos.

Reconocimiento del desapareamiento U:G por Msh2-Msh6

Como se ha descrito hasta ahora, los mecanismos de procesamiento de los desapareamientos U:G que implican la actividad de UNG explican principalmente las mutaciones en pares G:C. Sin embargo, en la reacción de SHM también se producen mutaciones en pares A:T en los genes de inmunoglobulinas. Según el modelo actual estas mutaciones surgen principalmente gracias al reconocimiento del desapareamiento U:G por la ruta de MMR y posterior procesamiento. La primera demostración de la implicación de MMR en la generación de mutaciones en pares A:T, provienen de la observación de que los ratones deficientes para Msh2 y Msh6 (Phung et al., 1998; Rada et al., 1998; Wiesendanger et al., 2000) tienen muy reducido este tipo de mutaciones.

Aunque el mecanismo exacto es aún desconocido, se ha propuesto que el desapareamiento U:G es reconocido por el heterodímero Msh2-Msh6 (Wilson et al., 2005)(Fig. 2). A continuación la exonucleasa I elimina un segmento de ADN adyacente al desapareamiento

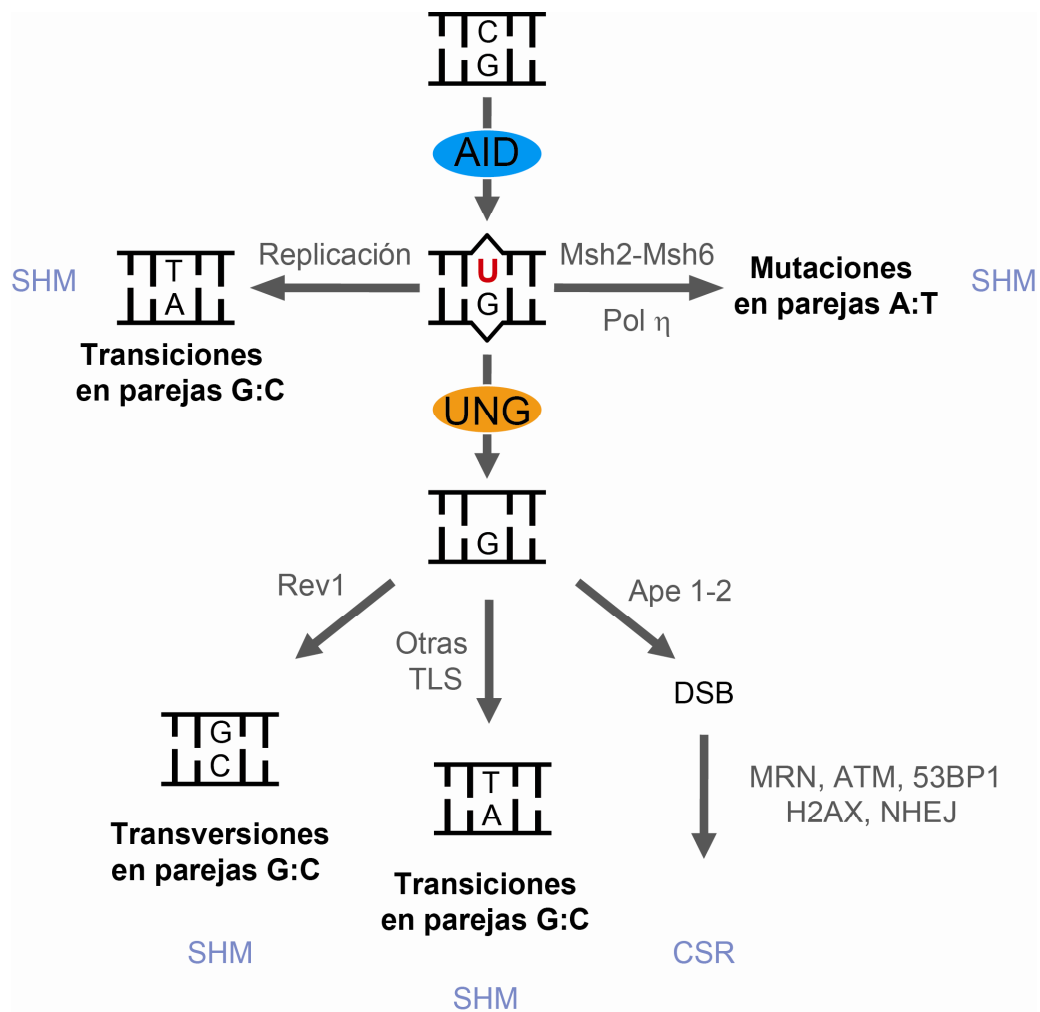


Figura 2. Mecanismos moleculares iniciados por AID

Representación del modelo de desaminación de citosinas por AID en el ADN. La desaminación de una citosina por AID genera un uracilo que se puede procesar de diversas formas, lo que determinará la aparición final de una mutación o de una rotura de doble cadena en el ADN (DSB). Si el uracilo no se repara antes de la replicación, se generarán transiciones en pares G:C. El desapareamiento U:G puede ser reconocido por el dímero Msh2-Msh6, paso inicial en el mecanismo de reparación de desapareamientos, que tiene como consecuencia final la introducción de mutaciones en pares A:T gracias a la ADN polimerasa η (Pol η). El uracilo puede ser reconocido también por UNG, que es una glicosilasa encargada de escindir este tipo de bases en el ADN y que forma parte de la ruta de reparación por escisión de base (BER). Se ha propuesto la participación de ADN polimerasas TLS en la resolución del sitio abásico generado, como Rev1 o la ADN polimerasa θ que estarían implicadas en la aparición de transversiones y transiciones en pares G:C. La generación del sitio abásico gracias a UNG también es un paso fundamental en la reacción de CSR, puesto que genera el sustrato para que actúen las endonucleasas Ape1 o Ape2 originando DSBs que son reconocidos por factores de respuesta al daño en el ADN (MRN, ATM, H2AX, 53BP1) y que se unirán por recombinación de extremos no homólogos (NHEJ).

U:G. Finalmente, se ha propuesto que Msh2-Msh6 son capaces de interaccionar con la ADN polimerasa η , favoreciendo así que sea esta enzima quien sintetice el nuevo ADN con un patrón de errores característico de esta polimerasa en el que prevalecen las mutaciones en pares A:T (Delbos et al., 2005; Wilson et al., 2005; Zeng et al., 2001) (Fig. 2). También se ha descrito que UNG podría estar implicada en menor medida en la generación de este tipo de mutaciones (Rada et al., 2004).

AID y la reacción de cambio de isotipo

La CSR implica la formación de DSBs en el ADN de las regiones de *switch* implicadas en la recombinación. AID inicia la reacción de CSR mediante la desaminación de citosinas en esas regiones (Martomo et al., 2005; Maul et al., 2011; Schrader et al., 2005). El uracilo generado es reconocido y escindido por UNG. Este paso es fundamental en CSR, puesto que los ratones deficientes para UNG tienen el CSR severamente disminuido (Rada et al., 2002), de la misma forma que los pacientes humanos con mutaciones que inactivan esta proteína (Imai et al., 2003). Al igual que en SHM, ninguna de las otras tres ADN-glicosilasas puede sustituir a UNG. A continuación la endonucleasa APE1 o APE2 crean una rotura de cadena sencilla en el ADN en el sitio abásico generado por UNG. El modelo actual propone que dos roturas de cadena sencilla en sitios próximos de hebras opuestas del ADN dan lugar al DSB necesario para el CSR. La recombinación entre las dos regiones de *switch* se realiza gracias a la maquinaria de reparación de extremos no homólogos (Stavnezer et al., 2008). Las primeras moléculas en reconocer el DSB son los componentes del complejo Mre11, Rad50 y Nbs1 (MRN), que activan a ATM para que fosforile a distintas proteínas implicadas en la reparación, como 53BP1 y H2AX, cuya función es fundamental para una correcta CSR (Gostissa et al., 2011). Existen más proteínas implicadas en la resolución de los DSBs, todas ellas necesarias para que la CSR sea eficiente, como XRCC4, ligasa IV, cernunnos, Ku70 y Ku80 (Manis et al., 2002; Manis et al., 1998; Pan-Hammarstrom et al., 2005; Soulas-Sprauel et al., 2007).

Efectos adversos de la actividad de AID en células B

La mayoría de los linfomas humanos se originan a partir de células B maduras (células B de centro germinal) y se caracterizan por la presencia de translocaciones cromosómicas linfomagénicas entre el *locus* de inmunoglobulinas y un proto-oncogén, como puede ser c-myc en linfoma de Burkitt, Bcl-1 en linfoma de células del manto, Bcl-6 en linfoma difuso de células B grandes o Bcl-2 en linfoma folicular (Kuppers, 2005). Como consecuencia de esta translocación se desregula la expresión del proto-oncogén. Existen estudios *in vivo* que demuestran que esta desregulación es crítica en el desarrollo de la neoplasia (Adams et al., 1985).

Se ha demostrado que AID promueve la generación de translocaciones entre c-myc y el *locus* de inmunoglobulinas, que surgen espontáneamente en ratones transgénicos para IL6 o en

células B estimuladas *in vitro* (Ramiro et al., 2006; Ramiro et al., 2004). AID es la responsable de la generación de los DSBs tanto en IgH como en c-myc, necesarios para que se produzca la translocación (Robbiani et al., 2008). La aparición de translocaciones depende también de UNG, lo que indica que tanto CSR como las traslocaciones c-myc/IgH se inician a través de un mecanismo común (Ramiro et al., 2006) (Fig. 1). La relevancia de la función de AID en la generación de linfomas se ha demostrado en diferentes modelos *in vivo*, incluídos los plasmacitomas inducidos por IL6 (Ramiro et al., 2004) o pristano (Kovalchuk et al., 2007) o los modelos de linfomas mediante expresión de c-myc (Kotani et al., 2007) y Bcl-6 (Pasqualucci et al., 2008).

Asimismo se han encontrado mutaciones características de la actividad de AID en diversos proto-oncogenes como Bcl6, Pim1, Myc y Rhoh en tumores difusos de células grandes (Pasqualucci et al., 2001), aunque no se ha demostrado su contribución al proceso linfomagénico. Más recientemente se ha descrito que AID podría desaminar citosinas presentes en el 25% de los genes expresados en células B normales y que el patrón de genes mutados estaba alterado en ratones deficientes para UNG y Msh2 (Liu et al., 2008). Este trabajo fue la primera evidencia de que AID podría actuar sobre un gran número de genes en células B y sugería que el procesamiento posterior del desapareamiento introducido por AID a través de UNG y MMR podría tener cierta especificidad. Sin embargo, se desconoce qué mecanismos son responsables de que determinados genes sean reparados fielmente mientras que en otros el procesamiento de lugar a la generación de una mutación.

En resumen, existen diversas evidencias de que la actividad de AID no está absolutamente restringida a los genes de inmunoglobulinas y puede dar lugar a lesiones en el ADN (mutaciones y DSBs) en otras secuencias del genoma.

Mecanismos reguladores de la actividad de AID

Lo expuesto anteriormente sugiere que la actividad de AID debe estar estrechamente regulada para evitar la introducción de daños potencialmente oncogénicos en el genoma y simultáneamente permitir una correcta respuesta inmune. Existen evidencias de los siguientes mecanismos reguladores de la actividad de AID.

1. Regulación transcripcional. En condiciones normales la expresión de AID está restringida mayoritaria pero no exclusivamente (como se describirá en el apartado siguiente) a células B

activadas. La inducción de AID está normalmente asociada a vías de señalización que intervienen en la activación de células B en el centro germinal, como las iniciadas por Interleuquina 4 (IL4), lipopolisacárido (LPS) o ligando de CD40 (CD40L) (Dedeoglu et al., 2004; Muramatsu et al., 1999; Zhou et al., 2003). Se han descrito una serie de elementos reguladores y factores de transcripción que controlan la expresión de AID, como E47, Id3, Id2, Pax5 y NF- κ B (Dedeoglu et al., 2004; Gonda et al., 2003; Ramiro et al., 2007; Sayegh et al., 2003; Shaffer et al., 2002; Tran et al., 2010). También se ha identificado una región muy conservada en el gen de AID que es fundamental para su transcripción *in vivo* (Crouch et al., 2007). Además, se ha descrito que los niveles de expresión de AID son críticos para su función, ya que se trata de una enzima haploinsuficiente y los animales heterocigotos para AID presentan la mitad de actividad que los *wild type* (Sernandez et al., 2008; Takizawa et al., 2008).

2. Regulación post-transcripcional. Los microARNs son moléculas de ARN pequeño no codificante que regulan la expresión génica afectando a la estabilidad o eficiencia de traducción de los ARNm a los que se unen. En nuestro laboratorio se ha identificado miR181b como un regulador de la CSR que disminuye la expresión de AID en células B activadas (de Yebenes et al., 2008). En otros trabajos se ha descrito que miR-155 también regula negativamente la expresión de AID y disminuye la frecuencia de traslocaciones prolinfomagénicas entre c-myc e IgH (Dorsett et al., 2008; Teng et al., 2008).

3. Regulación post-traduccional. La actividad de AID puede ser regulada mediante el control de su localización subcelular. La mayoría de la proteína se encuentra en el citoplasma gracias a la señal de exportación nuclear presente en su extremo carboxilo descrita anteriormente, siendo sólo activa cuando se facilita su entrada en el núcleo (McBride et al., 2004). Además, la actividad de AID aumenta cuando es fosforilada por la proteína kinasa A (PKA), lo que presuntamente facilita su interacción con la proteína de replicación A (RPA) y así su acceso al ADN de cadena sencilla (Basu et al., 2005; McBride et al., 2006; Pasqualucci et al., 2006). PKA se acumula específicamente en las regiones de *switch* implicadas en la reacción de CSR (Vuong et al., 2009). Finalmente, un trabajo reciente ha demostrado que la chaperona Hsp90 protege a AID de su degradación en el citoplasma (Orthwein et al., 2010).

4. Targeting de AID. Los mecanismos por los que AID actúa preferencialmente sobre los genes de inmunoglobulinas son desconocidos. Antes del descubrimiento de AID se había observado que la acumulación de mutaciones en la región variable de las inmunoglobulinas ocurría de forma preferencial en motivos WRCY/RGYW (Dorner et al., 1998; Rogozin and Kolchanov, 1992). Posteriormente se demostró mediante estudios bioquímicos que AID tiene preferencia

por mutar citosinas incluidas en motivos WRC (Bransteitter et al., 2004; Pham et al., 2003). Por lo tanto, existe una gran correlación entre la especificidad de AID *in vitro* y el patrón de mutaciones en *hotspots* observado en hipermutación somática, aunque no es idéntica. La posición situada en 3' de la C mutada juega un papel importante a la hora de definir un *hotspot* en la región variable de las inmunoglobulinas, mientras que no parece ser importante para definir la especificidad de AID *in vitro* (Rogozin and Diaz, 2004). La baja complejidad de este motivo no puede explicar la especificidad de AID para introducir mutaciones con alta frecuencia en el *locus* de inmunoglobulinas, lo que sugiere que la secuencia primaria de ADN no es el único determinante del *targeting* de esta enzima. En este sentido es importante mencionar que cuando la región variable de inmunoglobulinas se sustituye por secuencias heterólogas, éstas siguen siendo susceptibles a la función de AID.

El único requerimiento imprescindible descrito hasta la fecha para la actividad de AID es la activación transcripcional de la secuencia diana. Se ha descrito que la frecuencia de mutaciones en SHM es directamente proporcional a la tasa de transcripción en células B (Fukita et al., 1998). En CSR, el mecanismo principal para regular el cambio a un isotipo particular es la activación de la transcripción de la región de *switch* que precede a la cadena pesada de ese isotipo (Stavnezer et al., 2008). Se cree que la transcripción hace posible la generación transitoria de ADNcs, ya sea en forma de “burbujas de transcripción” (Ramiro et al., 2003) o de bucles del tipo “R-loops” (Yu et al., 2005) lo que permitiría la accesibilidad de AID a su sustrato. Estos mecanismos pueden explicar las mutaciones inducidas por AID en la hebra no transcrita del ADN, pero ambas hebras son mutadas por igual *in vivo*. Recientemente se ha propuesto que AID puede acceder a la hebra transcrita que está formando un híbrido con el ARN gracias a la interacción con el complejo del exosoma, encargado de degradar el ARN (Basu et al., 2011).. En dos estudios muy recientes se ha propuesto que la ARN polimerasa II (ARN polII) y Spt5, un factor de elongación de la transcripción, son capaces de reclutar a AID en sitios en donde la transcripción es más lenta (Pavri et al., 2010; Yamane et al., 2011). No obstante, a pesar de las evidencias que indican que la transcripción es indispensable para permitir la actividad de AID, se sabe que la simple expresión de un gen no es suficiente para convertirlo en diana de AID, puesto que muchos genes que se transcriben en el centro germinal no acumulan mutaciones (Liu et al., 2008; Shen et al., 1998). Además, los genes de inmunoglobulinas acumulan mutaciones con una frecuencia aproximada de 10^{-3} mutaciones por bp en cada división celular, mucho más alta que en cualquier otro gen expresado en células B y esta elevada tasa de mutación no se explica por un aumento similar en la intensidad

de transcripción (Di Noia and Neuberger, 2007). Por lo tanto, los mecanismos que regulan el *targeting* de AID siguen siendo objeto de intenso estudio en el campo.

Expresión y actividad de AID en células no B

La primera evidencia de que AID puede ejercer su función en células no B se obtuvo al comprobar que podía introducir mutaciones en un transgén transcripcionalmente activo cuando se sobreexpresaba en una línea celular de fibroblastos murinos, correlacionando la carga de mutaciones con la actividad transcripcional (Yoshikawa et al., 2002). Posteriormente se generaron ratones transgénicos para AID, en los que se observó que la sobreexpresión constitutiva de esta enzima tenía un efecto oncogénico sobre células T dando lugar a linfomas en los que se detectaron mutaciones puntuales en los genes del TCR y en c-myc. En estos animales se observaron también microadenomas en el epitelio de los bronquios de estos ratones (Okazaki et al., 2003).

Aunque originalmente la expresión de AID se consideraba exclusiva de células B activadas, algunos trabajos han descrito que puede detectarse en una serie de tejidos pluripotentes, como oocitos, células germinales embrionarias, células madre embrionarias (Morgan et al., 2004) y espermatocitos (Schreck et al., 2006). Se cree que la expresión de AID en estos tejidos puede contribuir de manera activa a la demetilación debido a su capacidad para convertir 5-metilcitosina en timidina (Bhutani et al., 2010; Morgan et al., 2004; Popp et al.; Rai et al., 2008). También se ha detectado expresión de AID en diversos tumores de origen no linfoide. Por ejemplo, al analizar líneas celulares epiteliales provenientes de pacientes con cáncer de mama se encontraron transcritos de AID (Babbage et al., 2006).

Los factores de transcripción de la familia NF- κ B tienen un papel muy importante en la regulación de la proliferación celular y la supervivencia y la activación constitutiva de esta vía se ha relacionado con la aparición de distintas neoplasias humanas (He and Karin, 2011; Luedde and Schwabe, 2011; Porta et al., 2011). En los últimos años se ha descrito una relación entre inflamación y expresión de AID a través de la vía de NF- κ B (Dedeoglu et al., 2004) y en líneas celulares derivadas de hepatocitos y epitelio gástrico se ha observado que citoquinas inflamatorias como TNF α pueden aumentar la expresión de AID a través de NF- κ B (Endo et al., 2007; Matsumoto et al., 2007). Además, se ha encontrado una correlación entre la presencia de AID y la aparición de mutaciones en p53 y c-Myc en hepatocarcinomas, cáncer gástrico y

carcinomas del ducto biliar (Komori et al., 2008; Kou et al., 2007; Matsumoto et al., 2007). Muy recientemente se ha descrito que los ratones deficientes en AID tienen una menor susceptibilidad a la generación de adenocarcinoma de colon mediado por inflamación (Takai et al., 2011).

Por lo tanto, AID es una enzima fundamental para una correcta respuesta inmune, cuya función es desaminar citosinas en el ADN que tiene como consecuencia la generación de mutaciones puntuales y DSBs en los genes de inmunoglobulinas. Además, está demostrado que puede ejercer su función sobre otros genes expresados en células B y en otros tipos celulares de forma aberrante. Esta tesis doctoral se ha centrado principalmente en el estudio del *targeting* de AID en un contexto de células no-B en dos niveles concretos: 1) especificidad de *locus* y 2) especificidad local de secuencia.

Objetivos

AID es una enzima que inicia la diversificación secundaria de anticuerpos a través de la desaminación de citosinas presentes en el ADN de los *loci* de inmunoglobulinas. Aunque esta actividad es fundamental para asegurar una respuesta inmune eficiente, AID también puede introducir lesiones potencialmente oncogénicas -tales como mutaciones o translocaciones cromosómicas- en otros sitios del genoma. Sin embargo, los mecanismos que regulan la especificidad funcional de AID son en gran medida desconocidos. El objetivo principal de esta tesis doctoral es tratar de entender estos mecanismos a través de dos abordajes concretos:

1. Estudio de la especificidad de *locus* de AID, mediante un análisis genómico de captura de genes susceptibles.
2. Estudio de la especificidad de secuencia de AID y de la contribución de los mecanismos de reparación de ADN, en particular de la glicosilasa UNG, en la definición de dicha especificidad.

Resultados

Desarrollo de un ensayo reportero para monitorizar las mutaciones introducidas por AID

AID es una desaminasa de citosinas cuya actividad es imprescindible para iniciar los dos mecanismos moleculares responsables de la diversificación secundaria de los anticuerpos, SHM y CSR (Muramatsu et al., 2000; Revy et al., 2000). Además de su función normal en células B, AID también promueve la aparición de translocaciones c-myc/IgH (Ramiro et al., 2006; Ramiro et al., 2004), introduce mutaciones en distintos genes expresados en células B (Liu et al., 2008; Pasqualucci et al., 2001) y su desregulación en células no B podría contribuir a la aparición de ciertos tipos de neoplasias (Komori et al., 2008; Kou et al., 2007; Matsumoto et al., 2007; Takai et al., 2011). Existen evidencias de que la generación de una mutación en genes distintos de las inmunoglobulinas a partir de la desaminación inicial introducida por AID podría depender del tipo de procesamiento (libre de errores o propenso a error) del desapareamiento U:G generado (Liu et al., 2008). Por ello, es de gran interés esclarecer la especificidad funcional de AID, así como la relación entre AID y la maquinaria de reparación.

Para tratar de responder a ambas cuestiones diseñamos en primer lugar un sistema fluorescente dual que permitiera monitorizar la generación de mutaciones inducidas por AID.

Generación de un casete reportero para monitorizar la actividad de AID

Para poder monitorizar la función de AID empleamos un sistema basado en la generación de células fluorescentes cuantificable por citometría de flujo que permitiera analizar un gran número de células de manera rápida y sencilla. Para ello se diseñó un casete reportero que consta de dos marcadores: OrangeSTOP y GFP, separados por un sitio de unión ribosomal (IRES, del inglés *Internal Ribosome Entry Site*). Para generar OrangeSTOP introdujimos una mutación puntual en el aminoácido Y77 de la proteína fluorescente Orange (Shaner et al., 2004) que sustituía el triplete TAC por TAG, creando un codón de parada de traducción (Fig. 3A) que daría lugar a una proteína trunca no fluorescente. Además, para conseguir que este nucleótido estuviese incluido en un *hotspot* para AID mutamos puntualmente el nucleótido siguiente, creando así una secuencia del tipo WRC (concretamente AGCT) y una mutación V78L. La proteína GFP sirve como marcador de transcripción del casete.

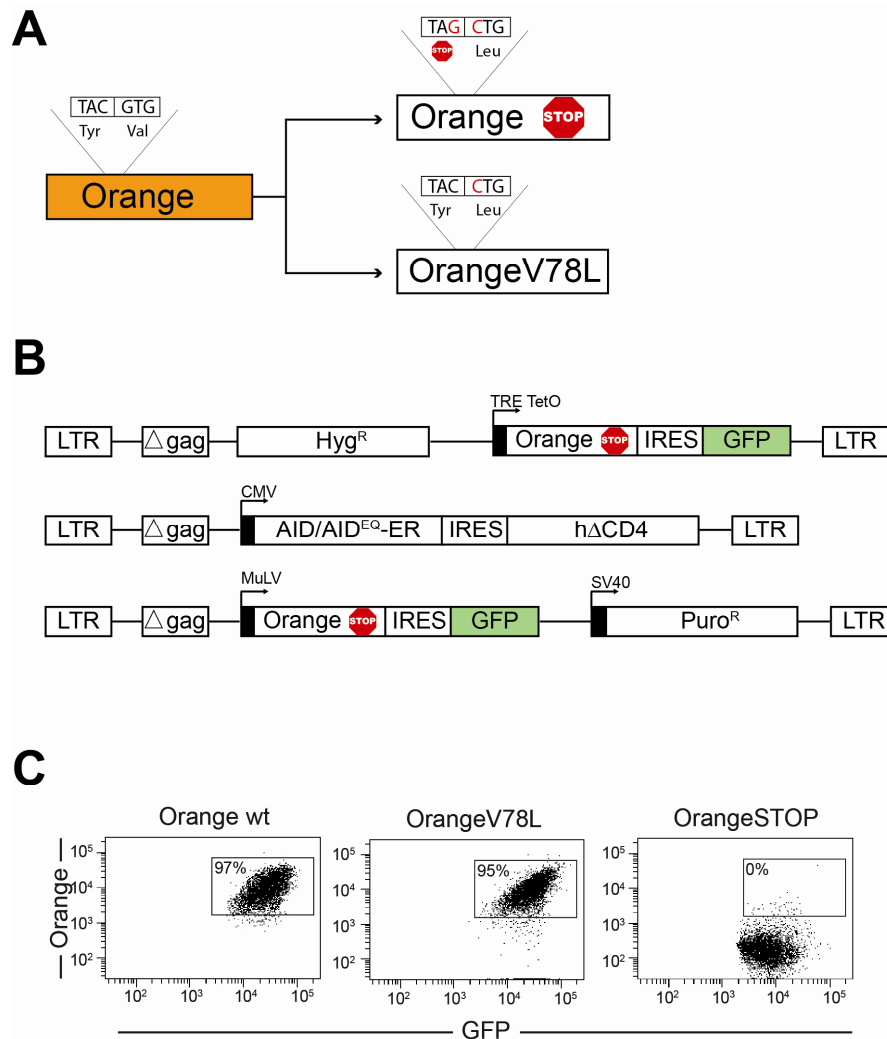


Figura 3: Diseño de una construcción para monitorizar mutaciones inducidas por AID

A) El casete OrangeSTOP se generó sustituyendo el codon TAC (Y77) por un codón de parada TAG. Para completar la generación del *hotspot* de hipermutación AGCT se introdujo el cambio adicional V78L. **B)** Vectores retrovirales empleados para monitorizar la actividad de AID. En primer lugar se representa el vector pREV-TRE con el casete reportero formado por OrangeSTOP y GFP separados por una secuencia IRES. Esta construcción se expresa bajo un promotor sensible a tetraciclina (TRE TetO). El vector incluye un gen de resistencia a higromicina (Hyg^R). En segundo lugar se representa el vector pQCXI-hΔCD4, con el que expresaremos AID o una forma catalíticamente inactiva AID^{E58Q} (EQ) fusionadas al dominio de unión a ligando del receptor de estrógenos (AID/EQ-ER). Contiene la secuencia de la proteína humana CD4 truncada. En tercer lugar se representa la construcción OrangeSTOP en el vector pMX-PIE, que contiene un gen de resistencia a puromicina (Puro^R). **C)** Análisis citométrico en el que se muestra el efecto de OrangeV78L y OrangeSTOP comparadas con la forma *wild type* de Orange (Orange wt).

La construcción OrangeSTOP se validó experimentalmente en un sistema NIH-3T3 *tet-off*. Para ello se transdujeron estas células con un vector que expresa el casete reportero OrangeSTOP-GFP bajo el control de un promotor sensible a tetraciclina (Fig. 3B, diagrama superior) y se comprobó mediante citometría de flujo el efecto de las mutaciones introducidas. Por una parte, se observó que la mutación en Y77 para introducir un codón de parada tiene como consecuencia el bloqueo en la formación de la proteína Orange funcional (Fig. 3C). También

comprobamos que la mutación V78L para crear un *hotspot* artificial no tiene ningún efecto sobre la fluorescencia de la proteína, puesto que se comporta igual que el control sin mutar (Fig. 3C).

Validación del casete reportero en un sistema inducible

Para comprobar si el casete diseñado permitía monitorizar las mutaciones introducidas por AID, empleamos un sistema de cotransducción retroviral en células NIH-3T3 para expresar conjuntamente el casete OrangeSTOP y AID.

AID es una proteína cuya localización es predominantemente citoplasmática, pero para acceder a su sustrato -citocinas en el ADN genómico- necesita traslocarse al núcleo. Para regular la actividad de AID a través de su localización subcelular, empleamos una versión de AID fusionada al dominio de unión a ligando del receptor de estrógenos (AID-ER). Esto permite que AID se movilice al núcleo en presencia de 4-hidroxitamoxifeno (OHT) (Doi et al., 2003). Como control negativo utilizamos una forma catalíticamente inactiva de AID, AID^{E58Q}, fusionada también a ER (EQ-ER).

Para expresar AID-ER o EQ-ER empleamos el vector pQCXI-hΔCD4. Esta construcción contiene la secuencia que codifica para la proteína CD4 humana truncada, que usamos como marcador de la eficiencia de transducción. La secuencia de AID-ER o EQ-ER están separadas del fragmento hΔCD4 por una secuencia IRES. (Fig. 3B, panel central). A estas construcciones las llamaremos AID-ER y EQ-ER. Para expresar el casete reportero OrangeSTOP-GFP usamos el vector retroviral pMX-PIE (Barreto et al., 2003) (Fig. 3B, panel inferior) que contiene un gen de resistencia a puromicina, lo que nos permite seleccionar con este antibiótico las células que han sido transducidas. A esta construcción la llamaremos OrangeSTOP-GFP para simplificar.

Para comprobar la eficiencia de la transducción simultánea de dos vectores se generaron sobrenadantes retrovirales mediante la cotransfección en células 293T de los vectores OrangeSTOP-GFP y AID-ER, con los que se cotransdujo la línea celular NIH-3T3. Tras seleccionar con puromicina observamos por citometría de flujo que un 91% de las células eran GFP⁺ y hΔCD4⁺, lo que indica que contienen ambas construcciones (Fig. 4A).

Para testar el sistema de monitorización de mutaciones añadimos OHT al medio de cultivo a una concentración de 1μM y analizamos la aparición de células Orange⁺. La figura 4B muestra un análisis de citometría representativo después de 7 días de cultivo en el que se observa la generación de células Orange⁺ en presencia de AID-ER, pero no cuando se expresa la forma

catalíticamente inactiva EQ-ER o cuando no se añade OHT al medio (no mostrado). La aparición de células Orange⁺ se puede detectar desde el día dos después de añadir OHT en células transducidas con AID-ER (Fig. 4C) y aumenta en días sucesivos de cultivo alcanzando un máximo a día 13. Estos datos indican que el casete reportero OrangeSTOP-GFP permite detectar la aparición de mutaciones, y que éstas son dependientes de la actividad de AID.

Para evaluar los niveles de AID en células NIH-3T3 transducidas con AID-ER en comparación con sus niveles fisiológicos de expresión se aislaron células B de bazo de ratones *wild type* y se estimularon con LPS e IL4. La estimulación *ex-vivo* de células B con LPS e IL4 es un sistema ampliamente utilizado que mimetiza la reacción de centro germinal, en cuanto a que se produce un incremento de la expresión de AID y el cambio de isotipo a IgG1. Se analizó la

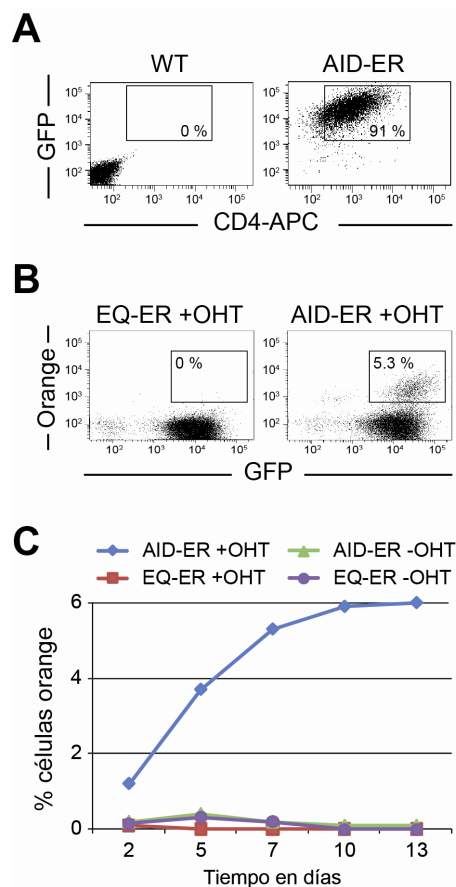


Figura 4: La construcción OrangeSTOP permite monitorizar las mutaciones inducidas por AID

A) Análisis por citometría de flujo de los marcadores GFP (construcción OrangeSTOP-GFP) y CD4 (construcción AID-ER) de células NIH-3T3 48 horas después de su cotransducción con los vectores pMX-PIE-OrangeSTOP y pQCXI-AID-ER hΔCD4. **B)** Análisis por citometría de flujo de células NIH-3T3 transducidas con OrangeSTOP-GFP y EQ-ER o AID-ER y cultivadas durante siete días en presencia de OHT. **C)** Experimento representativo en el que se muestra la cinética de aparición de células Orange⁺ en las células infectadas con el casete reportero y AID-ER cultivadas con OHT (rombos azules) o sin OHT (triángulos verdes), o EQ-ER con OHT (cuadrados rojos) o sin OHT (círculos morados).

cantidad de ARN mensajero (ARNm) por PCR cuantitativa a tiempo real (RT-qPCR) en ambos tipos de células y se observó un incremento de 6,5 veces en el sistema de cotransducción retroviral sobre la cantidad expresada en células B (Fig. 5A).

Para analizar la cantidad de AID en el núcleo se realizó una inmunofluorescencia en células NIH-3T3 cotransducidas con AID-ER y OrangeSTOP-GFP y cultivadas en presencia de OHT 1 μ M durante 5 días. De este modo observamos que en estas condiciones aproximadamente el 3% de la proteína AID total se localiza en el núcleo (Fig 5B). En células B está descrito que la

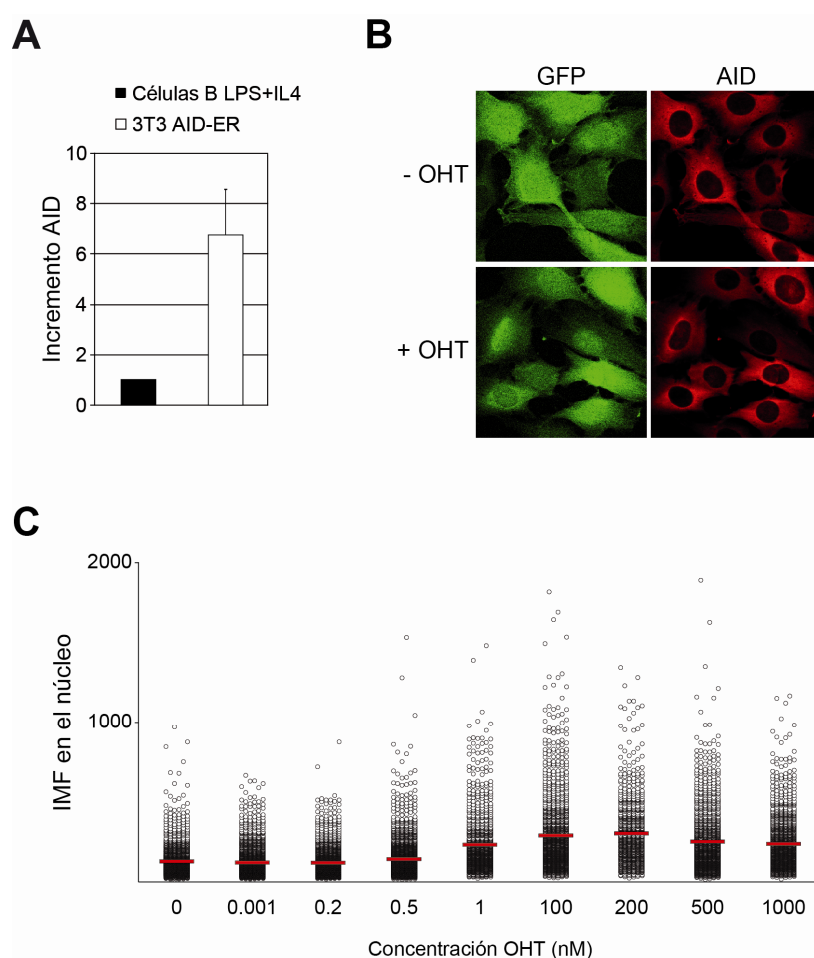


Figura 5: Análisis de la expresión de AID en células NIH-3T3 infectadas con AID-ER

A) RT-qPCR de AID en células NIH-3T3 infectadas retroviralmente con AID-ER y el casete reportero Orange-GFP seleccionadas durante 48 horas con puomicina (barra blanca) y de células B primarias de bazo estimuladas con LPS e IL4 durante 3 días (barra negra). La expresión de AID se normalizó frente a la de GAPDH y se representa el incremento respecto a la cantidad de AID en células B. La barra de error representa la desviación estándar (n=2). **B)** Análisis por inmunofluorescencia de células NIH-3T3 transducidas con AID-ER y el casete reportero Orange-GFP, cultivadas en presencia o ausencia de OHT durante 5 días, y teñidas con un anticuerpo anti-ER. **C)** Análisis mediante microscopía *high-throughput* de las células preparadas como se ha descrito en (A), cultivadas sin OHT o con concentraciones crecientes de OHT desde 0,001 nM hasta 1000 nM. Se analizaron al menos 400 células por cada concentración de OHT y se midió la intensidad media de fluorescencia (IMF) en el núcleo. Cada círculo representa un núcleo, la barra roja representa la media en cada caso.

cantidad de AID en el núcleo es de un 6% aproximadamente respecto a la cantidad total de AID de la célula, por lo que podemos estimar que la cantidad de AID en el núcleo en nuestro sistema es aproximadamente 3 veces mayor a la observada a niveles fisiológicos de expresión. También se realizó una cuantificación de la cantidad de AID en el núcleo de las células NIH-3T3 cotransducidas en presencia de distintas concentraciones de OHT en las mismas células y midiendo la intensidad de señal de AID en el núcleo por microscopía *high-throughput*. A concentraciones de OHT inferiores a 1nM no vemos ningún efecto en la movilización de AID al núcleo respecto al control sin OHT. Sin embargo, a 1nM se detecta un leve incremento en la intensidad nuclear media de AID (142U a 0,5nM, 232U a 1nM). La concentración de OHT empleada en las condiciones de cultivo descritas es de 1μM y la intensidad media en el núcleo es de 239U (Fig. 5C), lo que representa un incremento de 1,8 veces respecto a la intensidad detectada sin OHT.

Podemos concluir por tanto que hemos generado un sistema inducible que nos permite monitorizar las mutaciones inducidas por AID mediante la aparición de células Orange⁺ por citometría de flujo.

Identificación de *loci* permisivos a la actividad de AID

La única condición imprescindible conocida hasta la fecha para que un gen sea diana de AID es que se encuentre transcripcionalmente activo (Ramiro et al., 2003; Yu et al., 2005). Sin embargo, no todos los genes que se transcriben son susceptibles a la actividad de AID (Liu et al., 2008). Para tratar de esclarecer la especificidad funcional de AID nuestro objetivo fue desarrollar un sistema experimental de “caza” de sitios transcripcionalmente activos que sean permisivos con la actividad de AID. Este sistema debe cumplir dos condiciones: 1) disponer de un sistema para monitorizar la actividad de AID y 2) permitir evaluar selectivamente a nivel genómico aquellos *loci* que se encuentren transcripcionalmente activos. La monitorización de la función se realizará utilizando el casete OrangeSTOP-GFP descrito en la parte I. La evaluación selectiva de *loci* transcripcionalmente activos se abordará introduciendo el casete reportero en las células diana en un formato sin promotor.

Identificación de *loci* transcripcionalmente activos

Con el objetivo de llevar a cabo los experimentos de “caza” de promotores a partir de una población celular lo más homogénea posible en cuanto a expresión de AID, se transdujeron células NIH-3T3 con AID-ER o EQ-ER y se obtuvieron clones por dilución límite. Se obtuvieron un total de 22 y 26 clones de AID-ER y EQ-ER respectivamente, de los cuales 15 en ambos casos eran hΔCD4⁺ (dato no mostrado). Para comprobar que estos clones se comportaban según lo observado en la parte I, se transdujeron retroviralmente con el casete reportero OrangeSTOP-GFP, se cultivaron en presencia de OHT y se analizó la aparición de revertientes por citometría de flujo. En la figura 6A se puede observar el análisis después de 7 días de estimulación con OHT de dos clones representativos: B9 (AID-ER) y B6 (EQ-ER). El clon B9 cultivado en presencia de OHT daba lugar a una población de revertientes de un 11,8%. Por el contrario, no se detectaron células Orange⁺ tras el cultivo del clon B6 en presencia de OHT, de acuerdo con la dependencia de AID para la generación de mutaciones. Observamos que en los clones AID-ER existía una correlación ($R^2=0,96$) entre el porcentaje de células Orange⁺ y la actividad transcripcional del casete reportero, medida por la intensidad de GFP (Fig. 6B). Este resultado valida la utilización de clones para la monitorización de mutaciones inducidas por AID. Para los experimentos de “caza” de *loci* transcripcionalmente activos descritos a continuación, empleamos el clon B9 (en adelante, AID-ER) y el clon B6 (en adelante, EQ-ER).

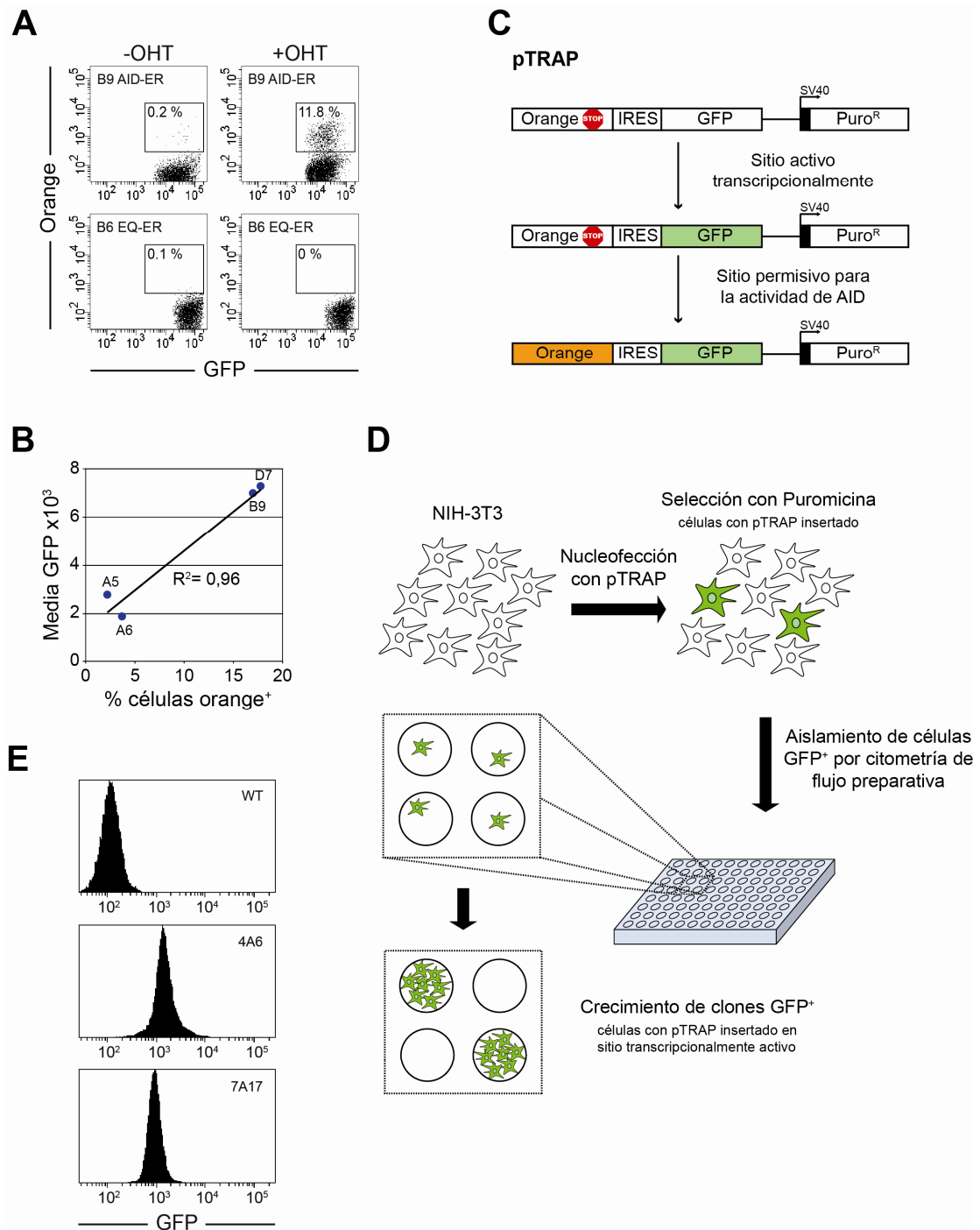


Figura 6: Estrategia experimental para la identificación de secuencias susceptibles a la actividad de AID

A) Análisis citométrico de clones representativos obtenidos por dilución límite a partir de células NIH-3T3 transducidas con AID-ER o EQ-ER y posteriormente con el casete OrangeSTOP-GFP y cultivadas durante 7 días en presencia o ausencia de OHT. **B)** Correlación entre porcentaje de revertientes y niveles de transcripción del transgén OrangeSTOP. Se representa el porcentaje de células Orange⁺ generadas por distintos clones AID-ER infectados con el casete OrangeSTOP-GFP y cultivados en presencia de OHT durante 15 días. La intensidad de expresión de GFP medida por citometría de flujo en cada clon individual indica el nivel transcripcional del transgén OrangeSTOP. **C)** Para cazar loci transcripcionalmente activos se diseñó la construcción pTRAP, que contiene el casete reportero OrangeSTOP-GFP sin promotor y un gen de resistencia a puomicina expresado bajo un promotor SV40. Sólo en el caso de integrarse en un sitio transcripcionalmente activo detectaremos expresión de GFP. Si el *locus* es accesible a la actividad de AID podremos monitorizarla mediante la aparición de células Orange⁺. **D)** Esquema de un experimento de “caza” de promotores transcripcionalmente activos. La construcción pTRAP se introduce por nucleofección en células NIH-3T3 y se seleccionan con puomicina durante 48 horas. A continuación las células GFP⁺ se aíslan por citometría de flujo preparativa de células individuales para obtener subclones. **E)** Análisis por citometría de flujo de dos subclones representativos en los que se ha cazado un sitio de inserción transcripcionalmente activo (GFP⁺).

Para realizar estos experimentos se creó una construcción no viral formada por el casete reportero OrangeSTOP-GFP sin promotor y el gen de resistencia a puromicina bajo un promotor propio. A esta construcción la llamaremos pTRAP (Fig. 6C). pTRAP se utilizó para transfectar los clones AID-ER y EQ-ER con el fin de conseguir su inserción al azar en el genoma. Al no tener promotor pTRAP sólo se expresará en aquellas células en las que se inserte lo suficientemente cerca de un promotor activo, que se identificará por la expresión de GFP (Fig. 6C).

Para introducir la construcción pTRAP en células NIH-3T3 se procedió a su linearización con el enzima de restricción *AseI*, se nucleofectó en los clones AID-ER y EQ-ER y éstos se mantuvieron en selección con puromicina durante 48 horas. A continuación se seleccionaron por citometría de flujo preparativa células GFP⁺ en condiciones de dilución límite y se crecieron los subclones obtenidos (Fig. 6D). El análisis posterior por citometría de flujo (Fig. 6E) confirmó la expresión de GFP estable en un 51% de los clones obtenidos (Tabla 1). Por lo tanto, concluimos que se ha conseguido “cazar” promotores transcripcionalmente activos.

Se realizaron cinco experimentos distintos de “caza” de promotores, de los que se obtuvieron un total de 126 subclones AID-ER y 74 EQ-ER, de los cuales 68 y 34 respectivamente mantenían de forma estable la expresión de GFP (Tabla 1).

Análisis de los *loci* transcripcionalmente activos

A continuación quisimos analizar si los promotores “cazados” en los distintos clones generados eran permisivos a la actividad de AID. Para ello se cultivaron todos los subclones AID-ER y EQ-ER GFP⁺ con OHT y se analizó la aparición de células Orange⁺ por citometría de flujo. Los porcentajes de revertientes eran variables en distintos subclones desde un 0,7% hasta un 2,2% a día 7 de cultivo con OHT (Fig. 7A) y el 22% de los subclones AID-ER GFP⁺ daba lugar a una población de células Orange⁺ (Tabla 1, Fig. 7B). En ningún caso se detectó una población de revertientes en los subclones EQ-ER GFP⁺ (Tabla 1, Fig. 7A, B). En adelante, a los subclones AID-ER en los que detectamos células Orange⁺ los llamaremos clones revertientes y en los que no las detectamos, clones no revertientes.

Es interesante destacar que el porcentaje de células Orange⁺ en los clones revertientes no correlacionaba ($R^2=0,4$) con la intensidad de transcripción de pTRAP medida como intensidad de señal GFP (Fig. 7C), analizado a día 7 de cultivo con OHT. Este resultado parece indicar que un aumento en la tasa de transcripción no garantiza una mayor actividad de AID.

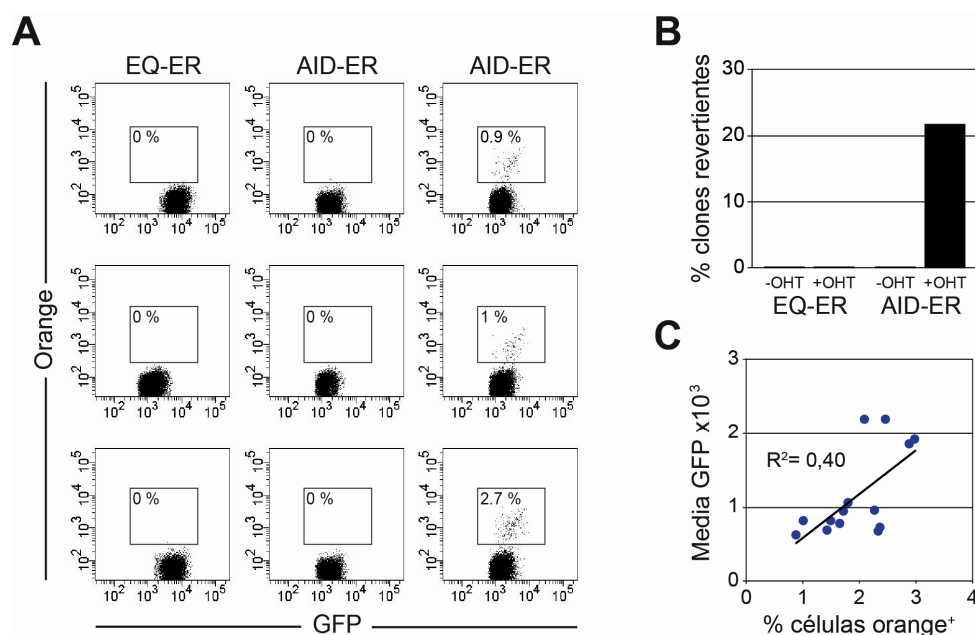


Figura 7: pTRAP permite evaluar de manera selectiva sitios transcripcionalmente activos y permisivos con la actividad de AID

Los subclones GFP⁺ generados como se ha explicado en la Figura 4 se cultivaron en presencia o ausencia de OHT. **A)** Análisis citométrico después de 7 días de cultivo con OHT de subclones representativos: 3 subclones EQ-ER (primera columna), 3 AID-ER Orange⁻ (segunda columna) y 3 AID-ER Orange⁺ (tercera columna). **B)** Porcentaje de subclones en los que se obtuvieron células Orange⁺ en las cuatro condiciones de cultivo (EQ-ER sin OHT, EQ-ER con OHT, AID-ER sin OHT, AID-ER con OHT). **C)** Ausencia de correlación entre el porcentaje de células Orange⁺ y la intensidad de transcripción medida como intensidad de señal GFP en los clones revertientes. Cada punto representa un clon.

Tabla 1: Número de subclones generados, GFP⁺ y Orange⁺ en los experimentos de “caza” de promotores

	B9 (AID-ER)	B6 (EQ-ER)
Clones generados	126	74
Clones GFP⁺	68	34
Clones Orange⁺	15	0

En base a estos resultados, podemos concluir que 1) hemos desarrollado un sistema de “caza” de promotores que nos permite evaluar genes transcripcionalmente activos y permisivos con la actividad de AID y 2) la transcripción del gen diana no es suficiente para garantizar la actividad de AID.

Determinación del sitio de inserción

Para determinar los sitios de inserción de pTRAP en los distintos subclones se realizó un ensayo de PCR inversa. Esta técnica permite amplificar fragmentos de ADN cuando uno de los extremos es desconocido y tiene tres pasos fundamentales: 1) Digestión del ADN con un enzima de restricción adecuado, 2) Ligación del producto de digestión generado y 3) Amplificación de la región de interés utilizando oligonucleótidos en orientación inversa. El primer paso se realizó digiriendo el ADN del clon a analizar con el enzima de restricción *Tsp509I*. La secuencia diana de *Tsp509I* se repite aproximadamente cada 500bp en el genoma de ratón y está presente en la posición 735 de OrangeSTOP. Tras la digestión obtendremos un fragmento de ADN en el que se incluye la parte correspondiente de pTRAP y un segmento del sitio en el que se ha insertado. A continuación se ligaron los extremos cohesivos generados y se amplificó la región de interés con dos parejas de oligonucleótidos en orientación inversa, que nos permiten realizar dos rondas de PCR, la segunda de ellas anidada (Fig. 8A). La construcción pTRAP se había transfectado linearizada con *AseI*, diana única en la construcción en el extremo 5' de OrangeSTOP, de manera que la secuencia comprendida entre las dianas *AseI* y *Tsp509I* se corresponderá con el sitio de inserción de pTRAP en el genoma celular. Los productos amplificados se separaron en un gel de agarosa y se clonaron las bandas de interés amplificadas. Como control, se digirió el ADN plasmídico de pTRAP con *Tsp509I*, se ligaron los extremos cohesivos y se amplificó el producto obtenido con los oligonucleótidos diseñados. Como resultado de esta amplificación se obtuvo un fragmento de 1,1 Kb, el tamaño esperado de la ligación entre las dianas *Tsp509I* 735 y 5562 (Fig. 8B, C).

Se analizaron 35 subclones AID-ER mediante esta técnica (Tabla 2). En tres de los subclones (4A10, 7A7 y 7A19), todos ellos revertientes, obtuvimos un producto de amplificación de aproximadamente 1,6 Kb (Fig. 8B). Al secuenciar estos productos se observó que el sitio en el que se había insertado pTRAP era idéntico en todos ellos (Fig 8C) y se correspondía con una secuencia de origen retroviral que no pudimos asignar a ningún cromosoma del genoma de ratón. Esta secuencia contenía la región codificante para la proteína *gag* de la envuelta vírica y parte de los elementos LTR. En adelante llamaremos RVS a este sitio de inserción encontrado. Dado que los tres subclones provienen de dos nucleofecciones distintas, podemos descartar que el resultado se deba a la amplificación de un clon individual en un experimento particular (Tabla 2).

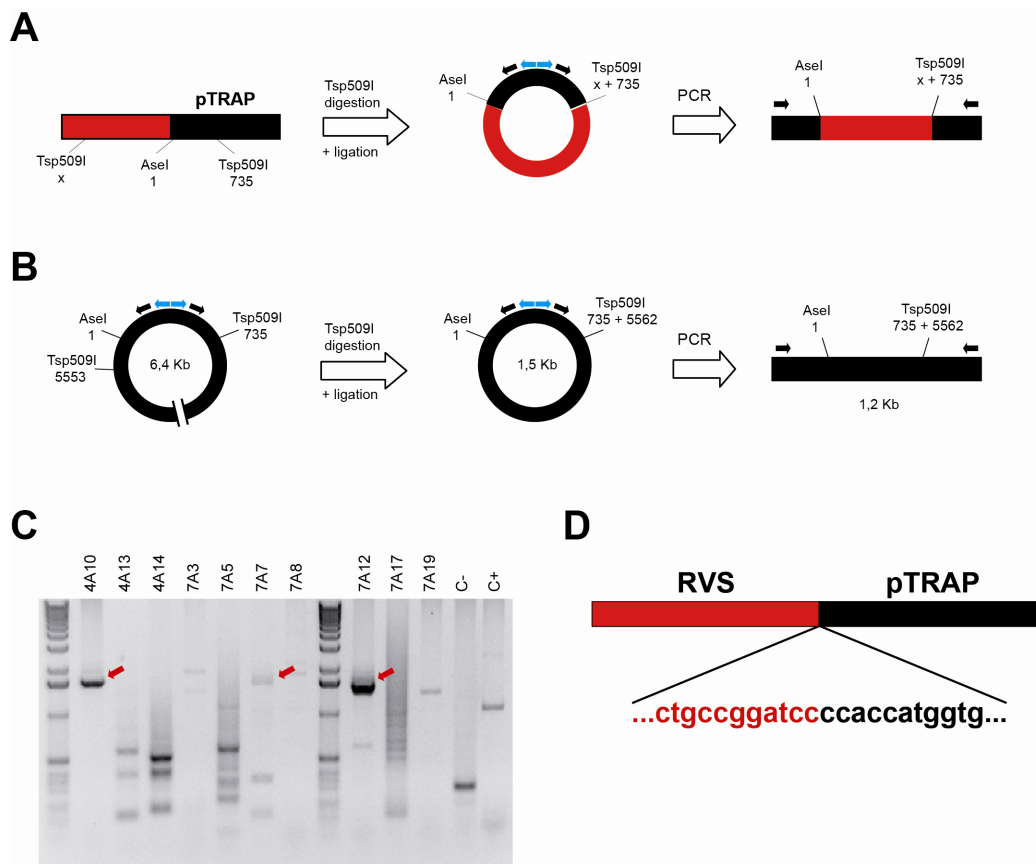


Figura 8: Determinación del sitio de inserción de pTRAP

A) Diseño del ensayo de PCR inversa (iPCR) utilizado para determinar los sitios de inserción de pTRAP. En el diagrama de la izquierda se representa pTRAP (rectángulo negro) insertado en un sitio desconocido (rectángulo rojo). Se indica la posición de la diana *Tsp509I* más proximal al sitio de inserción (posición 735), que está marcado por la posición de la diana *AseI*, utilizada para linearizar la construcción antes de su nucleofección. La digestión con *Tsp509I* generará un fragmento en el que se incluyen los primeros 735 bp comprendidos entre las dianas *AseI* y *Tsp509I* adyacente a un fragmento del locus de inserción flanqueado por otra diana *Tsp509I*, pero de tamaño desconocido. La ligación de este fragmento generado dará lugar a un ADN circular (diagrama central) susceptible de amplificarse con oligonucleótidos diseñados en orientaciones divergentes y representado por flechas. La amplificación se realiza en dos rondas (la primera, utilizando los oligonucleótidos representados como flechas azules, la segunda, con flechas negras). Como resultado de esta amplificación se obtendrá un producto que contiene un fragmento del sitio de inserción flanqueado por dianas *Tsp509I* y *AseI* (diagrama de la derecha). **B)** En el diagrama de la izquierda se representa pTRAP en su forma plasmídica, que se utiliza como control en los ensayos de iPCR. Se indican las dianas *Tsp509I* más proximal y distal a la posición de la diana *AseI*. La digestión con *Tsp509I* generará un fragmento en el que se incluye el ADN comprendido entre ambas dianas. La ligación de este fragmento generado dará lugar a un ADN circular de 1,5 Kb susceptible de amplificarse con los oligonucleótidos descritos en (A). El producto de amplificación tiene un tamaño de 1,2 Kb. **C)** Análisis en gel de agarosa al 1% de los productos obtenidos por iPCR en los clones indicados. Se utilizó pTRAP como amplificación control (C+) (ver (B)) y el clon B9 original como control negativo (C-). La flecha roja señala las bandas encontradas del mismo tamaño en los clones revertientes. **D)** Los productos de amplificación se secuenciaron y se observó que en todos los casos la secuencia comprendida entre las dianas *Tsp509I* y *AseI* era idéntica, con idéntico sitio de inserción, cuya secuencia se indica en la figura y se correspondía con una secuencia de origen retroviral que denominamos RVS.

Tabla 2: Subclones AID-ER generados en los experimentos de “caza” de promotores

^a Nombre del subclon, ^b subclones en los que se obtuvo (+) o no (-) el producto de amplificación de 1,6Kb mediante iPCR, ^c clones en los que se detectó (+) o no (-) la banda de 5Kb por *Southern blot*, ^d clones en los que se confirmó (+) o no (-) la inserción de pTRAP en RVS mediante PCR directa. El color naranja indica aquellos subclones en los que se obtuvieron células Orange⁺ después de cultivarlos en presencia de OHT. El color azul indica aquellos subclones en los que se confirmó que pTRAP estaba insertado en RVS pero no generaron células Orange⁺ después de cultivarlos con OHT. NT significa no testado.

Clon ^a	iPCR ^b	Southern ^c	PCR ^d
4A1	-	-	-
4A3	-	-	-
4A4	-	-	-
4A5	-	-	-
4A6	-	+	+
4A7	-	-	-
4A8	-	-	-
4A9	-	-	-
4A10	+	+	+
4A11	-	+	+
4A12	-	-	-
4A13	-	+	+
4A14	-	+	+
4A15	-	+	+
7A1	-	-	-
7A3	-	+	+
7A4	-	-	-
7A5	-	+	+
7A7	+	+	+
7A8	-	+	+
7A9	-	-	-
7A12	-	+	+
7A15	-	-	-
7A16	-	-	-
7A17	-	+	+
7A19	+	+	+
7A20	-	-	-
7A21	-	-	-
7A22	-	-	-
7A24	-	-	-
7A25	-	-	-
7A27	-	-	-
7A29	-	-	-
7A33	-	-	-

Clon ^a	iPCR ^b	Southern ^c	PCR ^d
7A34	-	-	-
8A1	NT	-	-
8A2	NT	-	-
8A4	NT	+	+
8A8	NT	-	-
8A10	NT	+	+
8A12	NT	-	-
8A19	NT	-	-
8A21	NT	-	-
8A23	NT	-	-
8A25	NT	-	-
8A27	NT	-	-
8A32	NT	-	-
8A33	NT	-	-
8A34	NT	+	+
10A1	NT	-	-
10A2	NT	-	-
10A3	NT	-	-
10A4	NT	-	-
10A6	NT	-	-
10A7	NT	-	-
10A9	NT	-	-
10A10	NT	-	-
10A11	NT	-	-
11A1	NT	+	+
11A3	NT	-	-
11A4	NT	-	-
11A5	NT	-	-
11A6	NT	+	+
11A8	NT	-	-
11A10	NT	-	-
11A11	NT	-	-
11A14	NT	-	-
11A20	NT	+	+

Para comprobar los datos obtenidos mediante la técnica de PCR inversa y realizar un análisis exhaustivo de los 68 subclones AID-ER y los 34 EQ-ER en los que hemos “cazado” un sitio transcripcionalmente activo se analizaron todos los subclones mediante *southern blot* utilizando como sonda un fragmento de la secuencia de GFP (Fig. 9A). Sorprendentemente, en los 15 clones AID-ER revertientes y en cuatro de los 53 no revertientes (4A11, 4A15, 11A1, 11A6) se detectó una banda de aproximadamente 5Kb (Fig 9B, panel izquierdo, Tabla 2). En los subclones EQ-ER también se encontró un sitio de inserción preferencial, aunque el fragmento de restricción detectado tenía un tamaño distinto al obtenido en los subclones AID-ER (Fig. 9B, panel derecho, Tabla 3).

Tabla 3: Subclones EQ-ER generados en los experimentos de “caza” de promotores

^a Nombre del subclon, ^b Ausencia (-) de células Orange⁺ después de cultivar los subclones con OHT, ^c subclones en los que se detectó (+) o no (-) la banda de 4Kb por *southern blot*.

Clon ^a	Orange ^{+b}	Southern ^c	Clon ^a	Orange ^{+b}	Southern ^c
4E1	-	-	7E9	-	-
4E3	-	-	7E11	-	-
4E4	-	+	7E13	-	-
4E5	-	-	7E16	-	+
4E6	-	-	7E22	-	+
4E7	-	-	7E23	-	-
4E8	-	-	7E26	-	+
4E9	-	-	8E1	-	-
4E10	-	-	8E2	-	-
4E11	-	-	8E3	-	-
4E13	-	+	8E4	-	+
4E14	-	+	8E5	-	+
4E15	-	+	8E6	-	+
4E17	-	+	8E7	-	-
7E1	-	+	8E8	-	-
7E4	-	+	8E10	-	-
7E6	-	+	8E11	-	+

Para confirmar mediante otra técnica esta inserción preferencial en RVS, se amplificó por PCR convencional la región de unión RVS-OrangeSTOP en los subclones AID-ER (Fig. 9A, C). Sólo se obtuvo producto de amplificación en aquéllos en los que se había detectado la banda de restricción de 5Kb por *southern blot* (Tabla 2). La secuenciación de estos productos confirmó que en todos los casos, la secuencia de inserción situada en 5' de pTRAP era RVS. Estos datos indican que existe un sitio preferencial de inserción de pTRAP en el genoma de NIH-3T3 y que todos los clones revertientes tienen la construcción integrada en ese mismo sitio.

Para determinar si el pTRAP insertado en RVS era efectivamente susceptible a la actividad de AID amplificamos por PCR la secuencia completa de OrangeSTOP específicamente insertada en el sitio retroviral identificado en el clon 4A10 y secuenciamos el producto obtenido. La frecuencia de mutaciones fue de $59,2 \times 10^{-4}$, con un patrón característico que indica actividad de AID, en el que predominan las mutaciones en pares GC (Fig. 9D). Esto confirma que el sitio de origen retroviral en el que se ha insertado pTRAP permite la actividad de AID.

En resumen, hemos podido establecer que pTRAP se ha insertado en la misma posición dentro de una secuencia retroviral en un 28% de los subclones AID-ER GFP⁺ en los que hemos “cazado” un sitio transcripcionalmente activo (19/68). De ellos, el 79% (15 de 19) son revertientes, es decir, permiten la actividad de AID, mientras que el 21% (4 de 19) restante no dio lugar a células Orange⁺. Los demás sitios transcripcionalmente activos que hemos conseguido capturar no permiten la actividad de AID en ningún caso (Fig. 10). En los subclones EQ-ER, el 44% presentan también una inserción preferencial, si bien se trata de un sitio distinto al descrito anteriormente y su secuencia no ha sido caracterizada (Tabla 3).

Concluimos que 1) existe un sitio preferencial de inserción, que se corresponde con una secuencia de origen retroviral que no hemos podido asignar a ningún cromosoma, 2) todos los clones revertientes tienen pTRAP insertado en esta secuencia retroviral y 3) Los clones revertientes dan lugar a porcentajes distintos de células Orange⁺, por lo que la susceptibilidad a la actividad de AID es distinta a pesar de tener el mismo sitio de inserción de pTRAP.

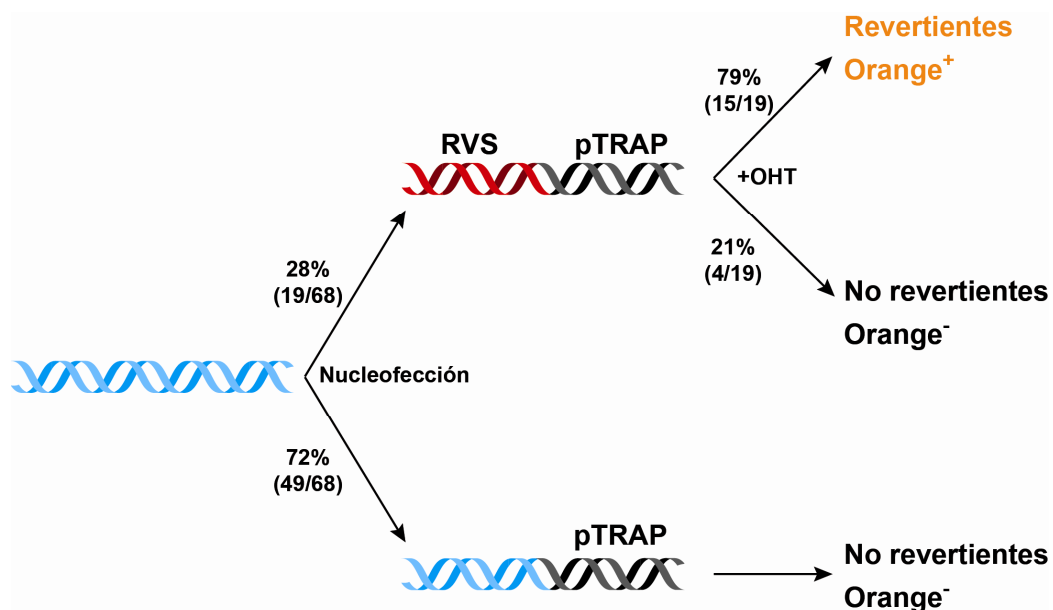


Figura 10: Diagrama resumen de los experimentos de caza de sitios susceptibles a la actividad de AID

Representación de la preferencia de inserción de pTRAP en células AID-ER. Se indican los porcentajes de clones para cada caso señalado y entre paréntesis los números de clones en cada caso / número total de clones analizados.

Análisis de la contribución de UNG a la especificidad de secuencia de la SHM

Como se ha descrito, AID es una desaminasa de citosinas que genera desapareamientos U:G en el ADN. Se ha propuesto que el uracilo generado puede ser replicado sin reparar, identificado por la maquinaria de reparación de desapareamientos MMR o bien reconocido por UNG (Di Noia and Neuberger, 2007). UNG es una glicosilasa de ADN cuya función consiste en escindir el uracilo, generando así un sitio abásico que puede ser procesado de distintas formas (Figs. 2, 11). UNG forma parte de la ruta de reparación por escisión de base (BER, del inglés *Base Excision Repair*). Sin embargo, en ratones deficientes para UNG el CSR está reducido drásticamente y el patrón de SHM está muy alterado, con una predominancia muy marcada de transiciones en pares G:C. Estos datos indican que UNG tiene un papel relevante en el contexto

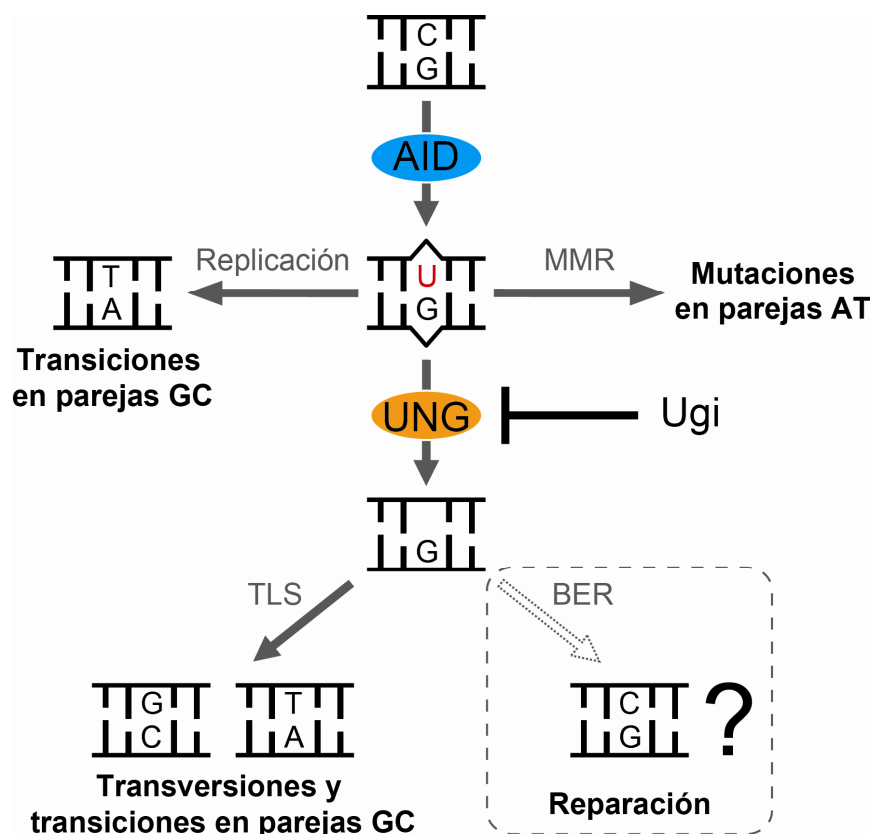


Figura 11: Sistema experimental para estudiar el balance entre AID y reparación de ADN

Esquema en el que se muestran las posibles vías que puede seguir una lesión introducida por AID para generar una mutación. AID desamina una citosina, generando un uracilo, que puede ser replicado sin reparar (transiciones en parejas CG), ser reconocido por la maquinaria de reparación de desapareamientos (MMR del inglés Mismatch Repair) (mutaciones en parejas AT), o ser escindido por UNG, generando un sitio abásico. Este puede ser reconocido por polimerasas TLS (del inglés *TransLesion Synthesis*) generándose transiciones o transversiones en parejas GC. No está claro si en este contexto UNG puede inducir la reparación del sitio abásico. Para comprobarlo, inhibiremos UNG con Ugi.

de la diversificación de anticuerpos en células B, presumiblemente a través de un procesamiento no canónico de los desapareamientos U:G inducidos por AID. Esta actividad aberrante del BER en el contexto de la diversificación secundaria de anticuerpos continúa siendo una de las grandes preguntas sin resolver en el campo.

Para tratar de esclarecer la función de UNG en la generación de mutaciones inducidas por AID llevamos a cabo una serie de abordajes de pérdida y ganancia de función combinados con un análisis de mutaciones mediante secuenciación masiva.

La inhibición de UNG bloquea CSR y la aparición de revertientes

Para comprobar el efecto de la inhibición de UNG en la actividad de AID utilizamos Ugi, una proteína producida por el bacteriófago PSB-2 (Di Noia and Neuberger, 2002). Está descrito que la expresión de Ugi en células B de pollo (DT40) modifica el patrón de mutaciones en la región V λ (Di Noia and Neuberger, 2002) y bloquea la reacción de CSR en la línea celular CH12 (Begum et al., 2004). Para comprobar el efecto de Ugi en la reacción de CSR de células B primarias, se aislaron células B de bazo, se estimularon en presencia de LPS e IL4, se transdujeron con Ugi en el vector pMX-PIE (Fig. 12A, panel superior) o con el vector vacío y se analizó el porcentaje de

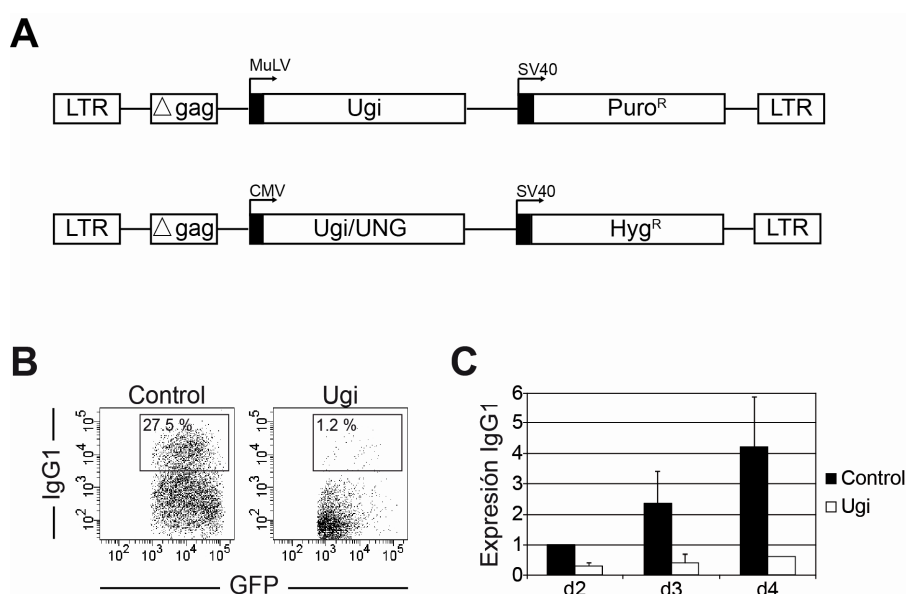


Figura 12: Ugi bloquea la reacción de CSR

A) Esquema de los vectores retrovirales empleados. En primer lugar está representado el vector pBABE-Hyg^R, con el que expresaremos un inhibidor de UNG producido por el fago PBS1 B subtilis (Ugi), o el propio UNG. Contiene la secuencia para el gen de resistencia a hygromicina (Hyg^R). En segundo lugar está el plásmido pMX-PIE, con el que expresaremos el casete reportero OrangeSTOP-GFP. En tercer lugar se representa el vector pQCXI-hΔCD4, con el que expresaremos AID-ER o EQ-ER. **B)** Estimulación de células B *ex vivo* activadas con LPS e IL4 y transducidas con Ugi, con UNG o el vector vacío durante cuatro días. Análisis citométrico a tres días de activación. Se analizó el porcentaje de células IgG1⁺ sobre células GFP⁺. **C)** Expresión de IgG1 normalizada a la expresión en el control tras dos días de estimulación (n=3-7. p (t de Student)<0.01 en todos los días.

células IgG1⁺ en cada caso. En las células que expresaban Ugi el CSR estaba severamente reducido (27,5% de células IgG1⁺ en el control frente a 1,2% con Ugi a día tres de estímulo, $p < 0.01$, t de Student) (Fig. 12B, C). Estos datos están de acuerdo con el efecto observado en ratones deficientes para UNG (Rada et al., 2002), por lo que podemos concluir que la expresión de Ugi en células primarias inhibe la actividad de UNG de manera eficiente.

Para evaluar el efecto de la inhibición de UNG en la aparición de células Orange⁺ en el sistema de monitorización OrangeSTOP, se transdujo Ugi en el vector retroviral pBABE-Hyg^R o el vector vacío como control (Fig. 12A, panel inferior). Esta construcción contiene un gen de resistencia a higromicina, por lo que se pueden seleccionar las células positivas con este antibiótico. El vector Ugi se cotransdujo con AID-ER y el casete reportero OrangeSTOP-GFP. Tras cultivar las células en medio con OHT, se observó que el porcentaje de células Orange⁺ estaba reducido más de un 90% en células que expresaban Ugi (5,7% control, 0,4% Ugi, Fig. 13A). El incremento en la aparición de revertientes a lo largo de los 20 días de cultivo con OHT es evidente en el control, llegando a un máximo de 12 veces tras 14 días de cultivo (Fig. 13), de acuerdo con los datos mostrados anteriormente (Fig. 4). Sin embargo, la inhibición de UNG a través de Ugi redujo la frecuencia de revertientes a todos los tiempos analizados ($p < 0.01$, t de Student) (Fig. 13B). Concluimos que la expresión de Ugi bloquea la generación de células Orange⁺ en el sistema reportero OrangeSTOP.

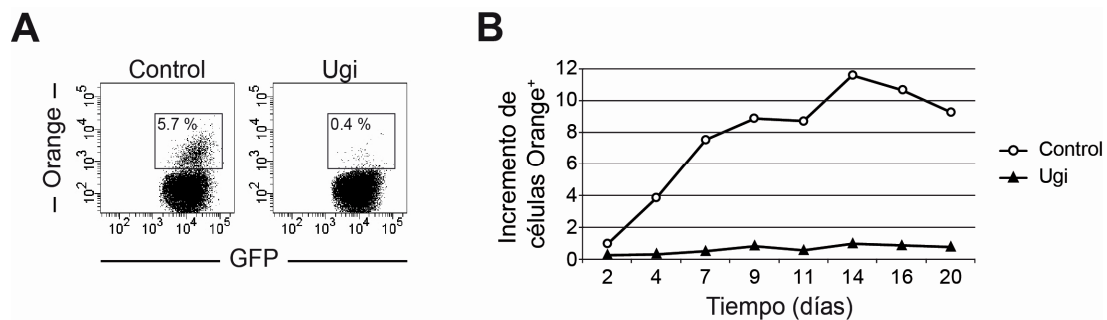


Figura 13: La expresión de Ugi bloquea la aparición de células Orange⁺

A) Análisis citométrico de células NIH-3T3 coinfectadas con AID-ER, el casete reportero OrangeSTOP-GFP y Ugi o el vector vacío tras siete días de cultivo en presencia de OHT. **B)** Cinética de aparición de células Orange⁺ en las mismas células cultivadas con OHT durante 20 días, normalizado al porcentaje de revertientes en el control a días dos (n=4-12, p (t de Student) < 0,01 a todos los días).

El codón de parada generado artificialmente en la proteína Orange es TAG (Fig. 3). Se ha demostrado que AID introduce preferencialmente mutaciones en pares G:C en genes expresados en líneas celulares no B (Yoshikawa et al., 2002). Las tres mutaciones que se pueden generar en la G son G→A, G→C y G→T, pero sólo las transversiones permiten revertir

el codón de parada y generar una proteína funcional, puesto que la transición G→A genera otro codón de parada TAA. Como se ha indicado anteriormente, la SHM en ratones deficientes para UNG se caracteriza por un incremento en la proporción de transiciones (Rada et al., 2002). Por lo tanto, la inhibición de la aparición de revertientes debido a la expresión de Ugi podría explicarse por un bloqueo en la generación de transversiones en pares G:C, lo que estaría de acuerdo con el fenotipo observado en ausencia de UNG. Esta hipótesis se aborda específicamente en el siguiente apartado.

La inhibición de UNG incrementa la frecuencia de mutaciones

Para caracterizar el tipo de mutaciones inducidas por AID en presencia y ausencia de Ugi, se transdujeron células NIH-3T3 con AID-ER, el casete reportero OrangeSTOP-GFP y Ugi o el vector vacío como control como se ha descrito en el apartado anterior. Tras 11 días de estimulación con OHT la secuencia de OrangeSTOP se amplificó mediante PCR y el producto generado se clonó. El análisis mediante secuenciación de un número suficientemente grande de colonias permite identificar las mutaciones que han surgido durante el cultivo en un número de células equivalente. Se secuenciaron un total de 101 y 76 colonias en el control y Ugi respectivamente, de las que se obtuvieron un total de 69 y 52 Kb de lectura (Fig. 14A). Se observó que la proporción de transiciones en pares GC estaba significativamente aumentada cuando se inhibe UNG (71% control, 98,3% Ugi, p (test de Fisher)= $2,8 \times 10^{-8}$) mientras que la proporción de transversiones en pares GC estaba disminuida (23% control, 0,8% Ugi, p (test de Fisher)= $3,5 \times 10^{-7}$) (Fig. 14B, C). Las mutaciones en pares AT son muy poco frecuentes tanto en células control como en Ugi⁺ (Fig. 14B, C) y la disminución observada en células Ugi⁺ no es significativa (6% control, 0,8% Ugi p (test de Fisher)=0,06). Estos datos están de acuerdo con el cambio en la proporción de mutaciones observado en ratones deficientes para UNG (Rada et al., 2002) y explica el bloqueo en la generación de revertientes que se ha descrito en el apartado anterior (Fig 13). Además de un cambio en el tipo de mutación introducido, también se observó que la frecuencia de transiciones en pares GC era considerablemente mayor ($7,1 \times 10^{-4}$ control, $22,5 \times 10^{-4}$ Ugi, $p=0,009$, t de Student), lo que causaba un aumento en la frecuencia total de mutaciones (10×10^{-4} control, $22,9 \times 10^{-4}$ Ugi, $p=0,014$, t de Student) (Fig. 14A, D). Estos datos demuestran por primera vez que el procesamiento por UNG de la desaminación introducida por AID no siempre tiene como resultado la introducción de una mutación, sino que UNG repara fielmente una proporción importante -alrededor de un 70% en este sistema experimental- de las lesiones inducidas por AID.

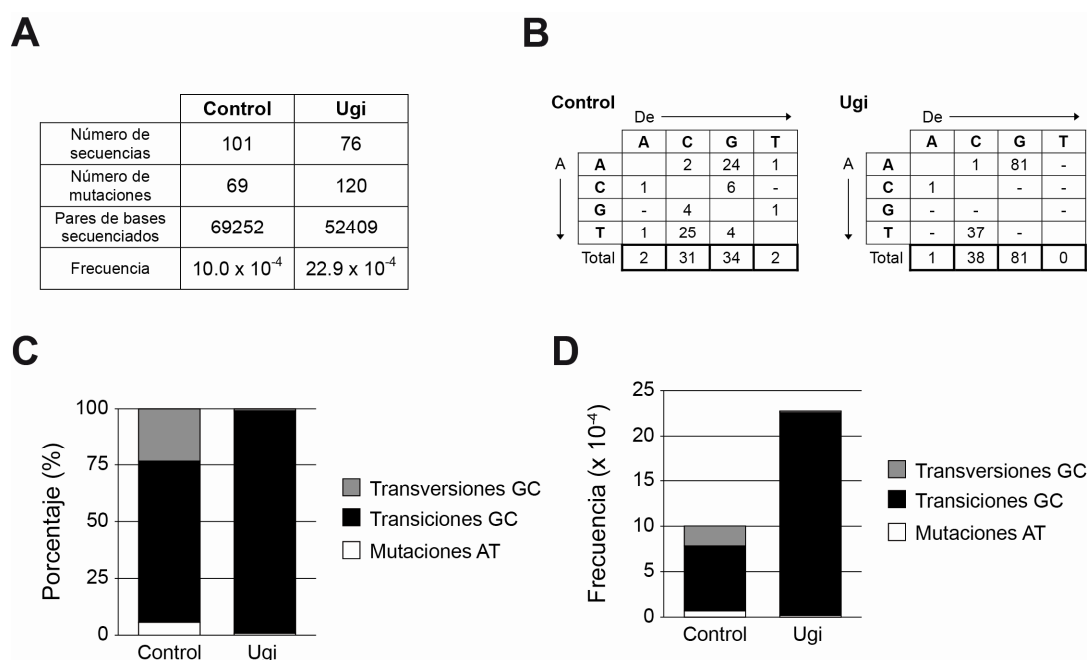


Figura 14: Ugi incrementa la frecuencia total de mutaciones en OrangeSTOP

Se amplificó, clonó y secuenció OrangeSTOP en células NIH-3T3 coinfectadas con AID-ER, el casete reportero OrangeSTOP-GFP y Ugi o el vector vacío cultivadas en presencia de OHT durante 11 días. **A)** Número de secuencias analizadas, mutaciones encontradas, pares de bases secuenciados y frecuencia de mutaciones en cada caso. **B)** Resumen de la distribución de mutaciones identificadas en las muestras control (panel superior) y Ugi (panel inferior). **C)** Contribución relativa de mutaciones en pares AT, transversiones y transiciones en pares GC. Los valores están expresados como porcentaje del total de mutaciones. (Control vs UGI transversiones y transiciones en GC, p (test de Fisher) $<0,0001$; mutaciones en AT, $p=0,06$). **D)** Frecuencias absolutas de transversiones y transiciones en pares G/C y mutaciones en pares AT, calculada como el número de cada tipo de mutación entre el total de pares de bases secuenciado. (Transversiones GC, p (t de Student) $=0,084$; transiciones GC, $p=0,0026$; mutaciones AT, $p=0,29$).

Análisis de las mutaciones inducidas por AID mediante secuenciación masiva

El número total de mutaciones secuenciadas en el experimento anterior es de 69 y 120 para células control y Ugi⁺ respectivamente (Fig. 14D). Si bien estos números son suficientes para analizar la frecuencia de mutaciones en términos globales, no lo son para obtener información detallada sobre el tipo de mutaciones y su distribución a lo largo de la secuencia y en los distintos *hotspots* de mutación descritos para AID. La secuenciación de un número suficiente de mutaciones para este fin mediante secuenciación tradicional (*Sanger*) sería muy costoso y requeriría mucho tiempo. Por ello se decidió poner a punto la detección de mutaciones inducidas por AID mediante secuenciación masiva de nueva generación (en adelante NGS, del inglés *Next Generation Sequencing*).

Para comprobar si era posible aplicar la tecnología NGS para el análisis de la actividad de AID se cotransdujeron células NIH-3T3 con AID-ER o EQ-ER, el casete reportero OrangeSTOP-GFP y Ugi o el vector vacío como se ha descrito anteriormente. Tras 11 días de estimulación con OHT se amplificó el gen Orange, se fraccionó el producto de amplificación y se creó una librería (ver materiales y métodos) en la que se introdujeron distintos códigos de barras para distinguir las secuencias procedentes de cada una de las muestras. Las librerías se secuenciaron en una plataforma Illumina y se obtuvieron lecturas de un tamaño de 36 bp de longitud cada una. El control EQ-ER se secuenció para poder estimar la tasa de error en el proceso de secuenciación, puesto que nunca se ha utilizado esta técnica para analizar las mutaciones inducidas por AID. Las secuencias obtenidas se alinearon y lo primero que se observó fue que la profundidad de lectura (número de veces que se ha leído cada nucleótido de OrangeSTOP) es tres órdenes de magnitud superior al obtenido en nuestros experimentos utilizando secuenciación de *Sanger* convencional (Fig. 15, tabla 4). La profundidad media obtenida en tres experimentos distintos fue de 89×10^3 lecturas por cada nucleótido en cada muestra secuenciada (Fig. 15, tabla 4). Es virtualmente imposible obtener la misma cantidad de información por secuenciación convencional. Se observó que la profundidad de lectura a lo largo de la secuencia de OrangeSTOP es irregular, con picos de 200×10^3 lecturas y valles de 3×10^3 lecturas (Fig. 15).

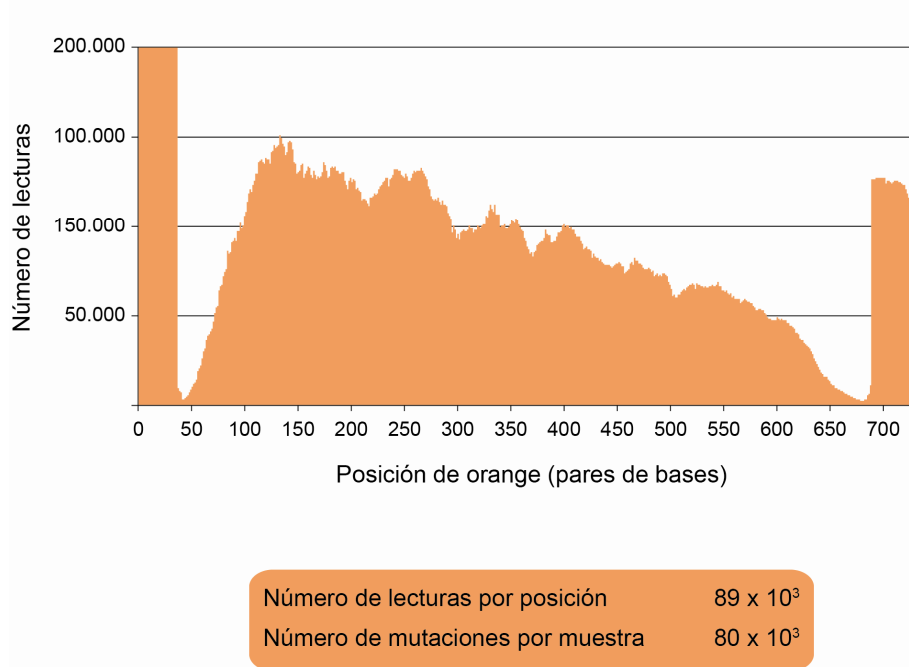


Figura 15: Profundidad de lectura en OrangeSTOP mediante secuenciación masiva (NGS)

Se amplificó el transgén OrangeSTOP de células NIH-3T3 coinfectadas con AID-ER o EQ-ER, el casete reportero OrangeSTOP-GFP y Ugi o el vector vacío y cultivadas en presencia de OHT durante 11 días. Tras fragmentar el producto de PCR y crear una librería (ver materiales y métodos), se secuenció por NGS. En la gráfica se representa el número de lecturas obtenidas por cada posición de OrangeSTOP. Debajo se indican el número medio de lecturas por cada posición y el número de mutaciones medio encontrado en cada muestra (n=3).

Mediante esta técnica se consiguieron detectar 22370 mutaciones en citosinas, frente las 31 detectadas por *Sanger* (Tabla 4).

Tabla 4: Secuenciación de OrangeSTOP por sanger o secuenciación masiva de nueva generación

^a Número total de citosinas presentes en OrangeSTOP, ^b Número medio de lecturas por cada citosina, ^c Número total de mutaciones en citosinas, ^d Frecuencia de mutación calculada como $c/(b \times a)$.

	Sanger	NGS
Número de posiciones^a	222	222
Número de lecturas (media)^b	101	87009
Número de mutaciones^c	31	22370
Frecuencia ($\times 10^{-4}$)^d	13,8	11,1

En primer lugar se comparó el patrón de mutaciones obtenido por *Sanger* y por NGS. Se analizó la acumulación de las mismas en el total de la secuencia y en *hotspots* en las muestras AID-ER en ausencia de Ugi. Como era de esperar, se observó que existe una acumulación significativa en la frecuencia de mutaciones detectada mediante secuenciación *Sanger* en *hotspots* de hipermutación WRC/GYW o WRCY/RGYW en células AID-ER⁺ (Fig. 16A). Esta misma tendencia se observó en los datos obtenidos mediante NGS (Total $9,86 \times 10^{-4}$, WRC $40,7 \times 10^{-4}$, WRCY $76,5 \times 10^{-4}$) (Fig. 16B). Por el contrario, las mutaciones detectadas en las células EQ-ER⁺ ocurren con una frecuencia baja, que no aumenta de forma específica en los *hotspots*, por lo que consideramos que esta frecuencia de mutaciones representa el fondo de la técnica de secuenciación (Fig. 16B). De estos datos, concluimos que es posible analizar mutaciones introducidas por AID mediante NGS. En adelante, las frecuencias asignadas a muestras AID-ER serán en realidad las resultantes de restar las frecuencias obtenidas en las muestras EQ-ER a sus correspondientes AID-ER.

A continuación se analizó la distribución de mutaciones en cada tipo particular de motivo WRCY/RGYW. En la figura 16C se muestra la frecuencia de mutaciones en las citosinas o guaninas de los distintos *hotspots* WRCY/RGYW. La línea roja representa la frecuencia media de mutaciones en pares GC a lo largo de toda la secuencia del transgén OrangeSTOP. Se encontraron grandes diferencias en la frecuencia de mutaciones detectada en cada uno de ellos, acumulándose preferentemente en las citosinas de AGCT y las guaninas de AGCA y AGCT, mientras que otras secuencias como TGCC, TACC ó GGCA no presentan frecuencias de mutaciones particularmente aumentadas (Fig. 16C). Para determinar si este patrón se detectaba también en la SHM canónica, se estimularon células B de bazo en presencia de LPS e

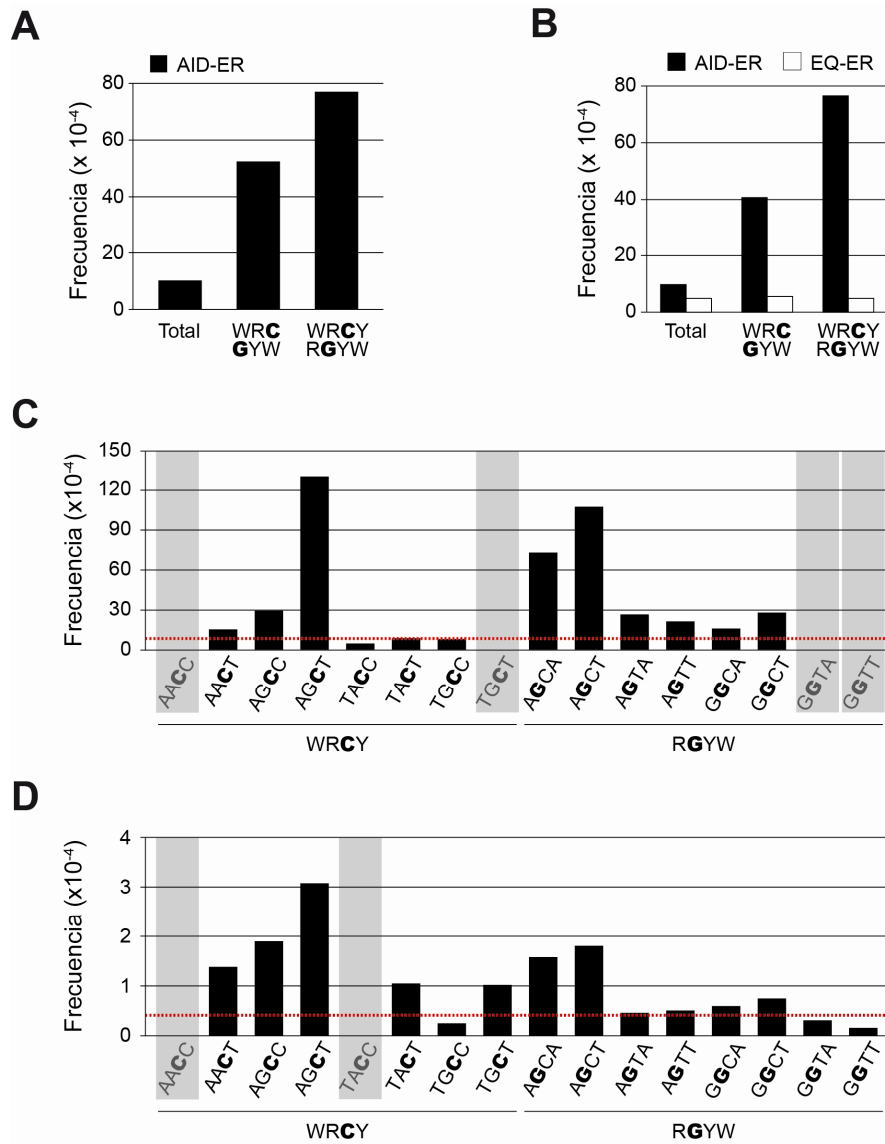


Figura 16: Análisis de las mutaciones inducidas por AID mediante NGS

Análisis del total de mutaciones detectadas en el transgén OrangeSTOP en células cotransducidas con AID-ER o EQ-ER, el casete reportero OrangeSTOP-GFP y Ugi o el vector vacío como control cultivadas en presencia de OHT durante 11 días. **A)** Frecuencia de mutaciones en las células AID-ER con el vector vacío secuenciadas mediante el método *sanger* y analizadas en el total de la secuencia de OrangeSTOP y en *hotspots* para AID (WRC/GYW o WRCY/RGYW, W=A o T, R=A o G, Y= C o T) **B)** Frecuencia de mutaciones en las células AID-ER (barras negras) o EQ-ER (barras blancas) con el vector vacío secuenciadas mediante NGS y analizadas como se describe en (A), calculada como la media de frecuencia en cada posición. En lo sucesivo, las frecuencias de mutaciones en AID-ER se representarán una vez sustraídas las frecuencias obtenidas en cada posición en células transducidas con EQ-ER. **C)** Frecuencia de mutaciones en cada uno de los motivos WRCY y RGYW en el transgén OrangeSTOP, calculada como se describe en (B). Las barras representan la media de tres experimentos independientes. La línea roja punteada representa la frecuencia media de mutación en pares GC detectada en el transgén OrangeSTOP. **D)** Análisis de mutaciones en S μ mediante NGS. Se estimularon células B *ex vivo* con LPS e IL4 durante 72 h procedentes de ratones AID^{+/+} y AID^{-/-}. Se amplificó la región S μ de dos animales diferentes de cada genotipo y se secuenció mediante NGS. La frecuencia de mutaciones se calculó como la media de los dos animales por cada posición, después de sustraer las frecuencias del AID^{-/-}, y se representa como se describe en (C).

IL4, se amplificó la región de cambio de isotipo μ (S μ) y se secuenció mediante NGS. La frecuencia detectada es bastante menor que en el caso de Orange, pero la distribución de mutaciones en los distintos tipos de *hotspots* es muy parecida en ambos casos (Fig. 16C, D). Las citosinas de AGCT y AGCC y las guaninas de AGCT son los sitios en los que más mutaciones se

acumulan. Estos datos indican que los distintos motivos WRCY/RGYW acumulan mutaciones de manera diferencial, sugiriendo que la especificidad de secuencia de la SHM es más restringida que la definida por esos *hotspots*. Además, el hecho de encontrar un patrón similar en la secuencia OrangeSTOP en células NIH-3T3 y en $\Sigma\mu$ de células B indica que esta especificidad no se restringe a un gen diana particular ni un tipo celular concreto.

Para determinar si el efecto de la inhibición de UNG en las mutaciones inducidas por AID es también detectable mediante NGS, se analizó la frecuencia y el tipo de mutaciones obtenidas cuando expresamos Ugi o el vector vacío. Se observó un incremento en la proporción de transiciones en pares GC (63,2% control, 99,8% Ugi) y una disminución en la proporción de transversiones GC (35% control, 0,2% Ugi) (Fig. 17A) muy similar al detectado por secuenciación *Sanger* (Fig. 14). La frecuencia en transiciones también estaba aumentada ($3,2 \times 10^{-4}$ control, 13×10^{-4} Ugi) lo que da cuenta de un aumento significativo en la frecuencia total de mutaciones (5×10^{-4} control, $13,1 \times 10^{-4}$ Ugi) (Fig. 17B) de forma equivalente al resultado obtenido por secuenciación tradicional. Estos datos validan de nuevo el método de NGS para secuenciar mutaciones inducidas por AID.

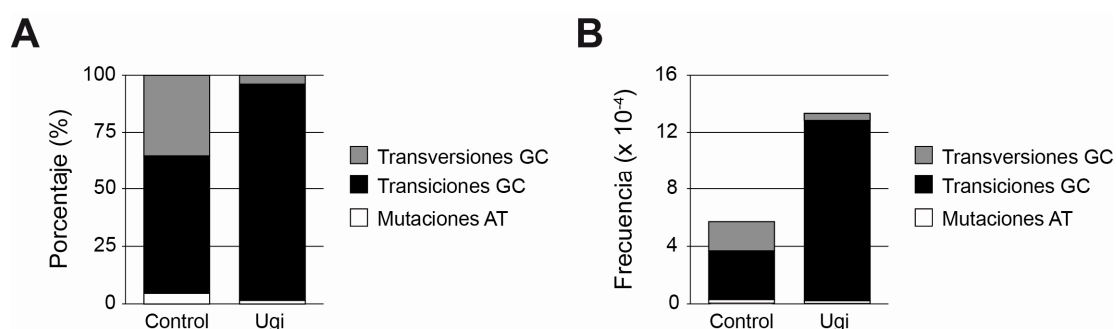


Figura 17: Patrón de mutaciones inducidas por AID

Se amplificó y secuenció mediante NGS el transgén OrangeSTOP de células NIH-3T3 coinfectadas con AID-ER o EQ-ER, el casete reportero OrangeSTOP-GFP y Ugi o el vector vacío y cultivadas en presencia de OHT durante 11 días. **A)** Contribución relativa de mutaciones en pares AT, transiciones y transversiones en pares GC en las células transducidas con Ugi o con el vector vacío. Los valores representan el porcentaje del total de mutaciones. **B)** Frecuencia absoluta de transversiones y transiciones en pares GC y mutaciones en pares AT calculada como la media de frecuencias en cada posición.

Según todo lo anterior concluimos que 1) podemos emplear NGS para estudiar la actividad de AID, 2) Los distintos *hotspots* WRCY/RGYW tienen diferente susceptibilidad a ser mutados por AID y esa preferencia se mantiene de forma independiente al tipo celular y a la secuencia diana y 3) la secuenciación NGS confirma que UNG repara parte de las lesiones inducidas por AID en niveles fisiológicos.

La sobreexpresión de UNG reduce el CSR y aumenta la aparición de revertientes

Los resultados anteriores demuestran por primera vez que UNG es capaz de reparar una fracción de las lesiones introducidas por AID. Esta observación abre la posibilidad de que los niveles fisiológicos de UNG no sean suficientes para hacer frente a la carga de lesiones U:G generadas por AID y ello dé lugar a una resolución mutagénica de las mismas.

Para evaluar si la cantidad de UNG es limitante en células B primarias, se estimularon con LPS e IL4, se transdujeron retroviralmente con UNG en el vector pMX-PIE o el vector vacío como control y se comprobó el efecto en el cambio de isotipo a IgG1. Se observó una reducción significativa ($p < 0,05$, t de Student en todos los tiempos analizados) de aproximadamente un 50% en la generación de células IgG1⁺ (Fig. 18).

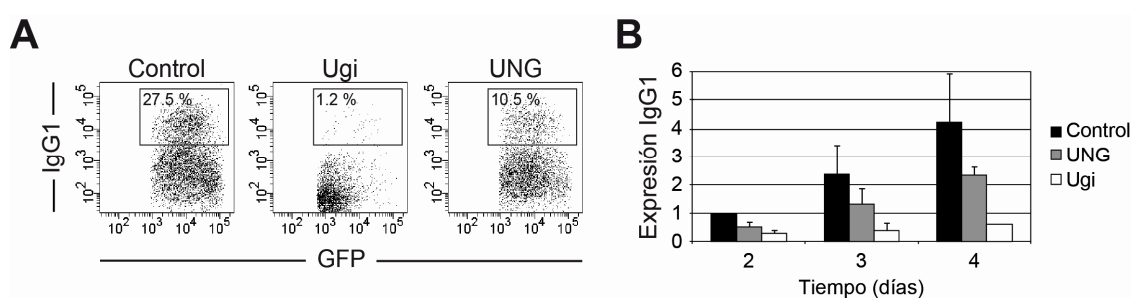


Figura 18: La sobreexpresión de UNG reduce el CSR

Se estimularon células B primarias de bazo con LPS e IL4, se transdujeron con Ugi, UNG o el vector vacío y se analizaron tras uno, dos o tres días de cultivo. **A)** Análisis citométrico después de tres días de activación. Se analizó el porcentaje de células IgG1⁺ sobre células GFP⁺. **B)** Expresión de IgG1 normalizada a la expresión en el control tras dos días de estimulación. (Ugi n=3-7, p (t de Student) < 0,01; UNG n=4-7, p < 0,05 en todos los días).

En segundo lugar analizamos el efecto de la sobreexpresión de UNG utilizando el ensayo de monitorización OrangeSTOP. Se cotransdujeron células NIH-3T3 con AID-ER o EQ-ER, el casete reportero OrangeSTOP-GFP y UNG, Ugi o el vector vacío y se estimularon con OHT durante 20 días. Se observó un incremento de entre un 40% y un 60% en la aparición de células Orange+ en las células transducidas con UNG respecto al vector vacío (5,7% control, 10,7% UNG, día 7 OHT) (Fig. 19).

Estos experimentos sugieren que la sobreexpresión de UNG tiene un efecto funcional en la resolución de las lesiones inducidas por AID, ya que disminuye la eficiencia del CSR en linfocitos B primarios y aumenta la frecuencia de revertientes en el ensayo OrangeSTOP. En los capítulos siguientes abordaremos los mecanismos moleculares responsables de estos efectos.

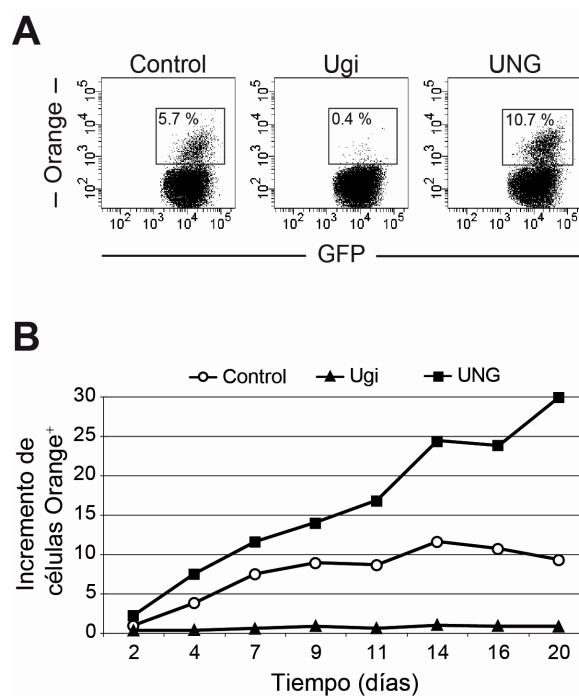


Figura 19: La sobreexpresión de UNG aumenta la aparición de células Orange⁺

Se cotransdujeron células NIH-3T3 con AID-ER, el casete reportero OrangeSTOP-GFP y Ugi, UNG o el vector vacío como control y se cultivaron en presencia de OHT. **A)** Análisis citométrico de las células control, Ugi y UNG tras siete días de cultivo con OHT. **B)** Cinética de aparición de células Orange⁺ en las mismas células cultivadas con OHT durante 20 días, normalizado al porcentaje de revertientes en el control a día dos. (Ugi vs control p (t Student) $<0,01$ a todos los días analizados; UNG vs control $p<0,05$ en todos los días, salvo día 7 y 9, donde $p=0,09$; $n=4-12$).

La sobreexpresión de UNG modifica el patrón de mutaciones inducidas por AID

Para analizar en mayor detalle el efecto de la sobreexpresión de UNG en las lesiones inducidas por AID, se amplificó y secuenció mediante NGS el transgén OrangeSTOP de células cotransducidas con AID-ER o EQ-ER, el casete reportero OrangeSTOP-GFP y UNG, Ugi o el vector vacío tras 11 días de cultivo con OHT. Se observó un incremento en la proporción de transversiones en pares GC en las células UNG⁺ con respecto a las células Ugi⁺ y al control (35% control, 0,2% Ugi, 65,2% UNG) y una disminución en la proporción de transiciones en pares GC (63,2% control, 99,8% Ugi, 34,6% UNG) (Fig. 20A). Además, la frecuencia absoluta de transversiones también estaba aumentada ($1,8 \times 10^{-4}$ control, $0,04 \times 10^{-4}$ Ugi, $2,8 \times 10^{-4}$ UNG) y la de transiciones disminuida ($3,2 \times 10^{-4}$ control, 13×10^{-4} Ugi, $1,5 \times 10^{-4}$ UNG), de modo que la frecuencia total de mutaciones no se encontraba alterada (Fig. 20B). La frecuencia de transversiones en el nucleótido G del codón de parada estaba incrementada en las células que sobreexpresan UNG respecto al control ($p=0.032$, t de Student) (Fig. 20C). Como se dijo antes,

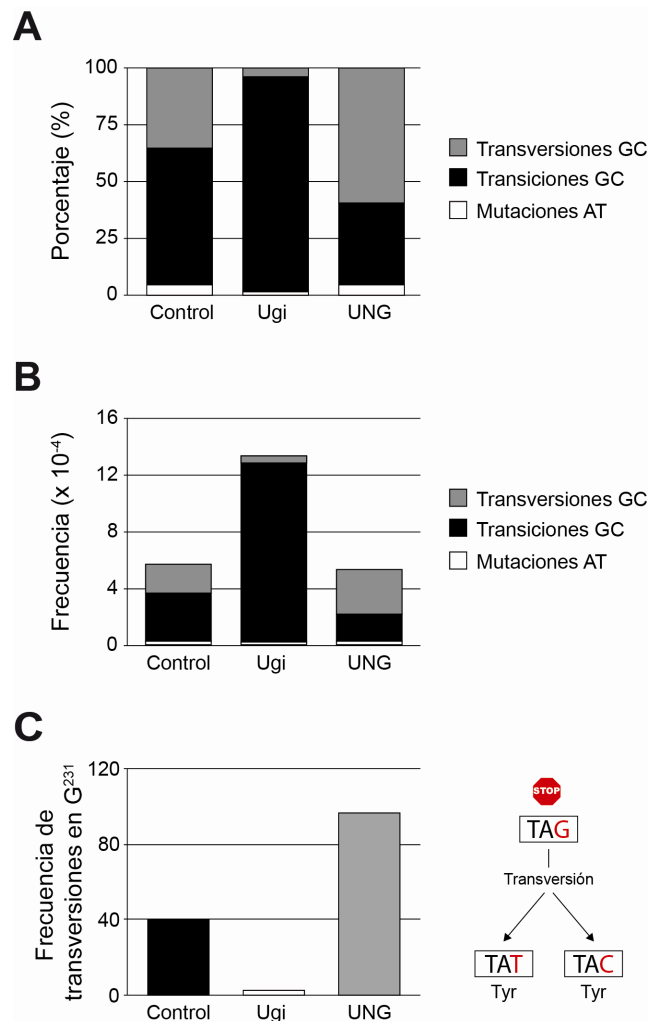


Figura 20: La sobreexpresión de UNG no provoca un aumento en la reparación, sino un cambio en el tipo de mutaciones

Análisis de mutaciones detectadas por NGS en el transgén OrangeSTOP en células cotransducidas con AID-ER o EQ-ER, el casete reportero OrangeSTOP-GFP y Ugi, UNG o el vector vacío como control. **A)** Contribución relativa de mutaciones en pares AT, transversiones y transiciones en pares GC. Los valores están expresados como porcentaje del total de mutaciones. **B)** Frecuencia absoluta de cada tipo de mutación, calculada como la media de frecuencias en cada posición. **C)** Análisis de mutaciones en el codón de parada generado en OrangeSTOP. Se representa la frecuencia de transversiones específicamente en el codón de parada en las células Ugi o UNG respecto al control. A la derecha se muestra los cambios en la posición analizada (G²³¹) que dan lugar a la reconstitución de fluorescencia (Orange⁺).

sólo las transversiones G→C o G→T en el codón de parada de OrangeSTOP permiten la formación de la proteína Orange fluorescente. Estos datos sugieren que la sobreexpresión de UNG provoca un aumento en la generación de células Orange⁺ debido a un incremento en la frecuencia de transversiones en el codón de parada. Concluimos que 1) la sobreexpresión de UNG no aumenta globalmente la reparación de mutaciones inducidas por AID y 2) UNG contribuye a la resolución final de la lesión inducida por AID ya que su sobreexpresión provoca un aumento de las transversiones y una disminución en las transiciones en pares GC.

UNG regula el balance entre reparación y mutación en *hotspots* de AID

Anteriormente hemos comprobado mediante NGS que AID introduce mutaciones preferencialmente en *hotspots* WRC/GYW y WRCY/RGYW y que cada tipo particular de *hotspot* acumula mutaciones de manera diferencial (Fig. 16). El hecho de que la sobreexpresión de UNG aumente el porcentaje de células revertientes Orange⁺ sin que exista un aumento en la frecuencia total de mutaciones, sugiere que la contribución de UNG a la generación de mutaciones a partir del procesamiento de lesiones U:G podría presentar cierta especificidad de secuencia. Para abordar esta cuestión se analizó la distribución de mutaciones en el transgén OrangeSTOP. Se analizó la frecuencia de transiciones y transversiones en pares G/C en los *hotspots* WRCY/RGYW individuales en las células control y en las que sobreexpresan Ugi o UNG. La sobreexpresión de Ugi provoca un aumento en las transiciones en todos los tipos de *hotspots* (Fig. 21A), mientras que la sobreexpresión de UNG implica mayoritariamente un aumento en la frecuencia de transversiones (Fig. 21B). Estos datos están de acuerdo con el análisis realizado en el total de la secuencia (Fig. 20B). Sin embargo, no todos los *hotspots* se comportan igual

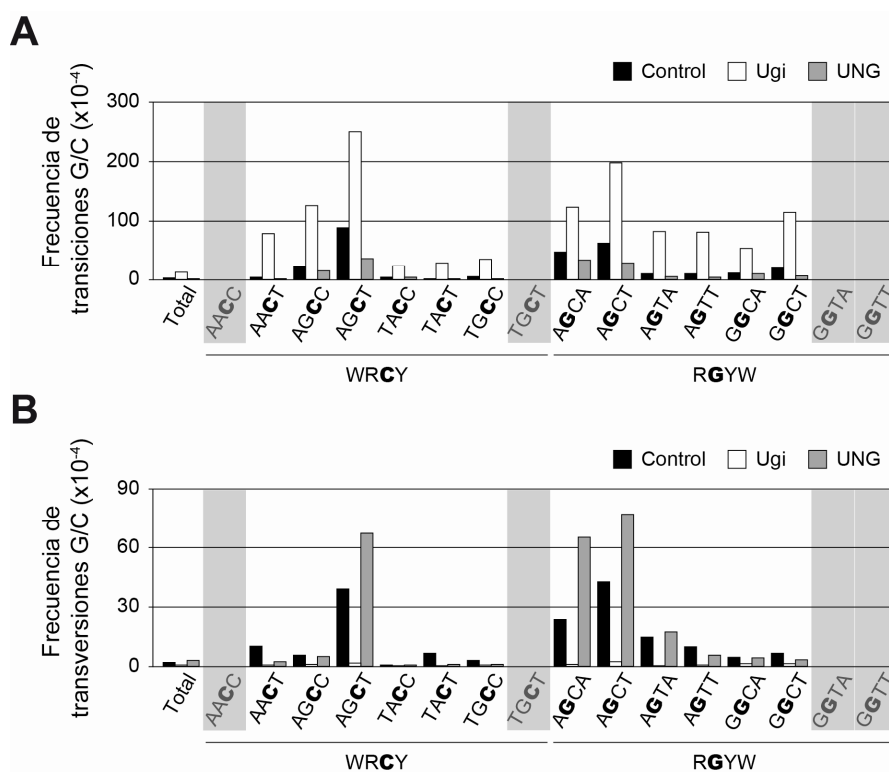


Figura 21: Análisis de mutaciones en *hotspots* para AID

Análisis de mutaciones detectadas por NGS en el transgén OrangeSTOP en células cotransducidas con AID-ER, el casete reportero OrangeSTOP-GFP y Ugi (barras blancas), UNG (barras grises) o el vector vacío (barras negras) como control. Se representa la frecuencia de transiciones (**A**) o transversiones (**B**) en pares GC en cada uno de los *hotspots* WRCY/RGYW en el transgén OrangeSTOP, calculada como la media de frecuencias en cada posición.

ante la variación de los niveles de UNG. Así, por ejemplo, la expresión de Ugi parece provocar un aumento más drástico en transiciones en AACT que en AGCA. Por otra parte, la sobreexpresión de UNG aumenta la frecuencia de transversiones en AGCT pero no en AGTA.

Para poder cuantificar este comportamiento diferencial, se compararon en primer lugar las frecuencias de transiciones en Ugi frente al control y en segundo lugar las transversiones en UNG frente al control. Hemos visto que al inhibir UNG con Ugi se incrementa la frecuencia total de mutaciones debido a un incremento en las transiciones en pares GC. Es lógico pensar que este incremento se producirá en posiciones en las que UNG normalmente repararía la lesión inducida por AID. Por lo tanto, para saber si UNG repara preferencialmente las lesiones inducidas por AID en algún tipo de secuencia particular, comparamos las frecuencias relativas de transiciones cuando expresamos Ugi respecto al control (Fig. 22A). De esta forma, se observó que si la desaminación ocurre en las citosinas de AACT, TACT o TGCC existe mayor predisposición a ser reparada fielmente (Fig. 22A). De manera inversa, para saber si existe

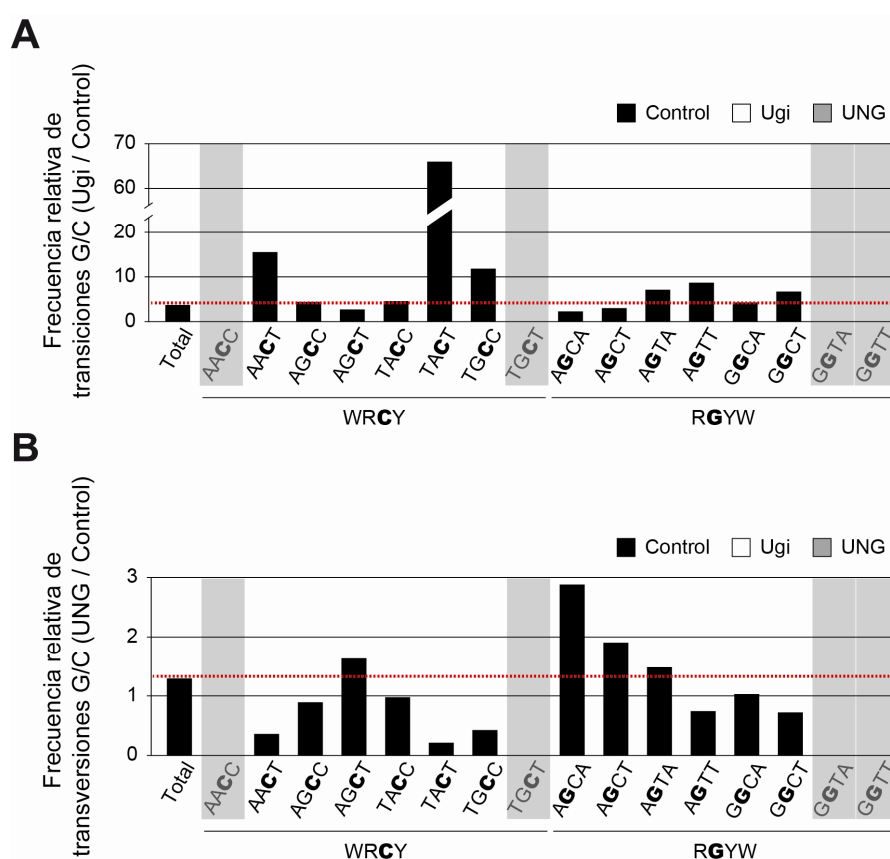


Figura 22: UNG repara o muta preferencialmente las lesiones inducidas por AID según el tipo de secuencia

Análisis de mutaciones en *hotspots* para AID detectadas por NGS en el transgén OrangeSTOP en células cotransducidas con AID-ER, el casete reportero OrangeSTOP-GFP y Ugi, UNG o el vector vacío como control. Se representa la frecuencia relativa de transiciones en pares GC **(A)** de Ugi respecto al control en el total de la secuencia y en cada uno de los *hotspots* WRCY/RGYW, o de transversiones en pares GC **(B)** de UNG respecto al control en los mismos tipos de secuencias.

alguna preferencia de secuencia en la reparación con tendencia a error mediada por UNG, comparamos las frecuencias relativas de las transversiones en pares GC cuando sobreexpresamos UNG respecto al control (Fig. 22B). Se observó que dos de los *hotspots* que más mutaciones acumulan (citosinas y guaninas de AGCT y guaninas de AGCA) tienen preferencia a ser reparados con tendencia a error, puesto que el ratio UNG/Control analizado en esas posiciones para transversiones en pares GC es mayor que la media en toda la secuencia de OrangeSTOP (Fig. 22B). Además, al comparar ambos análisis se comprobó que los ratios Ugi/control para transiciones en pares GC son una imagen especular de los ratios UNG/control para transversiones en pares GC (Fig. 22).

En resumen, podemos afirmar que 1) UNG repara una parte de las lesiones inducidas por AID, 2) UNG no se limita a escindir los uracilos generados por la actividad de AID, sino que contribuye a la resolución de estas lesiones y 3) UNG presenta cierta especificidad de secuencia que ayuda a definir el patrón de mutaciones inducidas por AID.

Discusión

Ensayo de revertientes y NGS para el análisis de la actividad de AID

En este trabajo se ha abordado la especificidad de locus y secuencia de la actividad de AID. Para ello se ha desarrollado un sistema de monitorización de mutaciones inducidas por AID basado en un casete reportero OrangeSTOP-GFP. Se escogió la proteína fluorescente Orange para este estudio porque su espectro de emisión permite su combinación con la de otras proteínas fluorescentes de uso muy extendido, como GFP. Anteriormente se habían desarrollado otros sistemas de monitorización de mutaciones basados en la aparición de señal fluorescente debida a la reversión de una mutación que trunca GFP (Bachl and Olsson, 1999; Yoshikawa et al., 2002). En nuestro caso, utilizamos un sistema fluorescente dual en el que la expresión de GFP permite seguir las células que han incorporado el casete reportero y podemos analizar sobre ellas la aparición de revertientes Orange⁺ generadas por las mutaciones introducidas por AID. De esta forma, la monitorización de la actividad de AID es más precisa, lo que será de gran utilidad para ensayos en células primarias, en las que la eficiencia de transducción retroviral es muy baja.

Se ha demostrado que AID es un enzima haploinsuficiente, por lo que sus niveles de expresión son determinantes para su actividad (Sernandez et al., 2008; Takizawa et al., 2008). Además, al sobreexpresar retroviralmente AID en células B activadas se incrementa la frecuencia de traslocaciones c-myc/IgH de manera notablemente mayor que un aumento proporcional a la cantidad de AID (Ramiro et al., 2006; Ramiro et al., 2004). También se ha propuesto que un aumento en los niveles de AID podría dar lugar a una pérdida de especificidad de *locus* o secuencia (Pasqualucci et al., 2008). Hemos estimado que a pesar de utilizar un sistema de transducción retroviral, en nuestros ensayos la cantidad de AID nuclear es tan sólo unas tres veces mayor que en células B activadas. Por otra parte, la comparación de células 3T3 transducidas *versus* células B muestra un espectro de mutaciones muy similar. Estos datos validan la utilización de este ensayo para el estudio de la actividad y especificidad de AID y su extrapolación a otros contextos, incluido el de células B activadas. Estos resultados, junto con observaciones anteriores (Sernandez et al., 2008; Takizawa et al., 2008) sugieren que aunque

el aumento de la expresión de AID tiene como resultado un impacto cuantitativo en su actividad, no altera sustancialmente su especificidad.

Es inesperado que sólo un 3% de la proteína se encuentra en el núcleo después de cultivar las células cotransducidas con OHT. La fusión del dominio ER a AID se ha descrito anteriormente como un sistema con el que se consigue una movilización al núcleo muy eficiente (Doi et al., 2003), aunque la regulación de la localización subcelular de AID es muy compleja y pequeñas diferencias experimentales pueden influir en la correcta traslocación de AID al núcleo. Por ejemplo, se sabe que la forma humana, que sólo difiere de la murina en dos posiciones de la secuencia de exportación nuclear se trasloca más eficientemente al núcleo (di Noia, comunicación personal). En nuestro caso, medimos la cantidad de AID nuclear mediante inmunofluorescencia, mientras que en el trabajo de Doi et al realizan un análisis mediante *western blot* tras una fragmentación subcelular, por lo que la comparación entre ambos estudios sólo puede ser aproximada. Además, los vectores de expresión utilizados no son los mismos y por tanto la cantidad de proteína expresada puede ser muy distinta, lo que puede dar cuenta de diferencias en la cantidad de AID nuclear. En cualquier caso, esta deficiente traslocación al núcleo representa una ventaja en nuestro sistema, puesto que nos permite realizar los experimentos en unas condiciones más parecidas a las fisiológicas.

En este trabajo se ha puesto a punto la detección de mutaciones inducidas por AID mediante NGS. La NGS es un método de secuenciación muy utilizado en los últimos años que ha permitido un gran avance en la detección de mutaciones asociadas a distintos tipos de neoplasias (Ngo et al., 2011; Puente et al., 2011) y ha abaratado y acelerado la secuenciación de genomas de muchos organismos (Li et al., 2010; Rohde et al., 2011). La mejora en los sistemas de NGS es constante, permitiendo cada vez la obtención de un mayor número de lecturas de alta calidad. En nuestro caso, hemos implementado un protocolo de NGS para detectar mutaciones introducidas por AID, de baja frecuencia y no clonales ya que ocurren de forma esporádica en cada célula individual del cultivo con una frecuencia de 10^{-5} a 10^{-3} mutaciones por par de base. Por este motivo, para el estudio de las mutaciones inducidas por AID es fundamental que el protocolo de secuenciación cumpla dos objetivos fundamentales: 1) debe permitir la obtención de un número muy alto de lecturas por cada posición y 2) la tasa de error del sistema de secuenciación debe ser menor que la frecuencia de mutaciones inducidas por AID. Respecto al primer requisito, existen distintos protocolos empleados para enriquecer partes específicas del genoma (Varela et al., 2011; Wang et al., 2011), aunque con estas técnicas se alcanzan en el mejor de los casos entre 200 y 300 lecturas por posición. Con nuestra aproximación, la distribución de lecturas obtenida es muy variable a lo largo del gen

secuenciado por razones que desconocemos, aunque parece estar relacionado con la secuencia primaria ya que amplicones distintos presentan patrones de profundidad característicos. En cualquier caso, el número de veces que se ha leído cada posición oscila entre 3×10^3 y 200×10^3 , varios órdenes de magnitud superior a la obtenida mediante técnicas de enriquecimiento. En cuanto al segundo requisito, el protocolo de secuenciación NGS empleado implica dos pasos distintos de amplificaciones por PCR que pueden ser una fuente de errores inespecíficos. Los resultados obtenidos en células en las que se expresó el mutante catalíticamente inactivo EQ-ER demostraron que la frecuencia de mutaciones inducidas por AID era detectable por NGS y que ésta está sensiblemente aumentada en secuencias definidas como *hotspots* de hipermutación. El hecho de que las lecturas obtenidas por NGS tengan una longitud reducida (36bp en nuestro caso), implica realizar un análisis de mutaciones únicamente por posición, e impide obtener información de secuencias completas del transgén, equivalentes a la obtenida mediante secuenciación por el método *Sanger*. A pesar de que esta limitación imposibilita el análisis de clones individuales, los datos obtenidos son perfectamente válidos para realizar el estudio de frecuencias y distribución de mutaciones como el que se ha realizado en este trabajo.

En definitiva, creemos que este tipo de abordaje va a ser extremadamente valioso para estudios futuros en el ámbito de la hipermutación somática.

Loci susceptibles a la actividad de AID

AID actúa preferencialmente sobre los genes de inmunoglobulinas, aunque se desconocen los mecanismos que determinan esta especificidad. En la segunda parte de este trabajo se ha desarrollado una estrategia con la que se pretendía abordar los mecanismos que regulan el *targeting* de AID. Existen trabajos en los que se ha detectado expresión de AID en células no B, como en espermatoцитos (Schreck et al., 2006), oocitos, células madre germinales y células madre embrionarias (Morgan et al., 2004) aunque su relevancia funcional en estos tejidos no está bien establecida. También se ha detectado expresión de AID en tumores no linfoides y se ha sugerido una conexión entre la actividad mutagénica de AID en estos tejidos y la aparición de este tipo de neoplasias (Babbage et al., 2006; Matsumoto et al., 2007; Takai et al., 2011). Por todo esto consideramos importante desarrollar un sistema en el que poder estudiar el *targeting* de AID en un contexto no B. En este proyecto empleamos la línea celular de fibroblastos murinos NIH-3T3.

Quisimos identificar de manera sistemática y a nivel genómico los *loci* susceptibles a la actividad de AID para posteriormente encontrar parámetros comunes que pudieran dar cuenta de esta susceptibilidad, tales como niveles de expresión, secuencias reguladoras, contenido de *hotspots* de hipermutación, etc. Para ello se diseñó una construcción sin promotor ni secuencias LTR (pTRAP) que nos permitiera en primer lugar limitar el estudio a *loci* transcripcionalmente activos, ya que la transcripción es el requerimiento mejor establecido para la actividad de AID. En nuestro ensayo, la resistencia a puomicina hace posible seleccionar las células nucleofectadas, y la expresión de GFP indica que la integración ha ocurrido en un sitio con actividad transcripcional endógena. En estos ensayos se nucleofectaron aproximadamente 5×10^6 células en cada experimento, lo que hace un total de 25×10^6 células analizadas. Creemos que este número podría haber sido suficiente para estimar a nivel genómico global la susceptibilidad a la actividad de AID de un número considerable de *loci* transcripcionalmente activos, aunque el resultado obtenido indica que la integración de pTRAP en el genoma no ocurre al azar (ver más adelante) y por lo tanto no se ha cumplido ese objetivo. En los experimentos de “caza” de promotores aproximadamente el 50% de los clones obtenidos tras seleccionar por citometría de flujo preparativa células GFP⁺ mantuvieron la expresión de GFP de forma estable. Se ha descrito que el ADN exógeno introducido en el genoma de mamíferos puede ser metilado, inhibiendo así su transcripción (Suzuki et al., 2007; Yang et al., 2007) por lo que es posible que pTRAP pueda ser reconocido como ADN exógeno, quizá en sitios particulares de inserción, dando lugar al silenciamiento del promotor bajo el cual se expresa, lo cual podría explicar la desaparición de GFP en la mitad de los clones obtenidos.

Nuestros análisis por Southern blot, iPCR y PCR convencional indican que la inserción de pTRAP en el genoma no es al azar, sino que ocurre de forma preferencial en sitios concretos del genoma. El 28% de los subclones AID-ER GFP⁺ analizados (19/68) presentaban el mismo sitio de inserción. Se descartó que estuviésemos amplificando el mismo clon porque los subclones revertientes provienen de cuatro experimentos distintos. También se observó que el 44% de los subclones EQ-ER tenían otro sitio de inserción común en todos ellos, distinto al detectado en los AID-ER. Aunque se ha descrito anteriormente que la integración de plásmidos tiene un sesgo hacia sitios transcripcionalmente activos (Suzuki 2010), y que existe una cierta tendencia por sitios frágiles (Merrihew et al., 1996; Rassool et al., 1991), nunca se había descrito una preferencia tan marcada por un sitio particular como la que muestran nuestros resultados. Los mecanismos de integración de ADN exógeno en el genoma no están claramente definidos, aunque se sabe que son distintos para vectores virales y no virales. En sistemas no virales

como el nuestro se ha propuesto que la integración ocurre gracias a procesos de reparación del ADN y recombinación (Haber, 1999; Van Dyck et al., 1999). En estudios previos en los que se había introducido un ADN exógeno no viral en distintos tipos celulares, se identificaron distintos sitios de integración a nivel de secuencia con características particulares – microhomologías, pequeñas deleciones e inserciones- que indican que este proceso se ha producido gracias a un mecanismo de recombinación no homóloga (NHEJ, del inglés *Non Homologous End Joining*) (Merrihew et al., 1996; Suzuki et al., 2010; Wurtele et al., 2003). En el sitio de integración estudiado no detectamos estas pequeñas diferencias, aunque no podemos descartar que en todos los sitios de integración estudiados se hayan generado las mismas deleciones o inserciones, generando por tanto un sitio de inserción idéntico en experimentos distintos.

Hemos identificado el sitio de inserción, en el caso de los subclones AID-ER y se corresponde con una secuencia de origen retroviral que hemos denominado RVS. Se ha descrito que aproximadamente el 40% del genoma de los mamíferos deriva de retroelementos, que son elementos genéticos móviles que contienen el gen de la transcriptasa reversa. De ellos, la cuarta parte son retrovirus endógenos (Waterston et al., 2002) y se cree que provienen evolutivamente de infecciones por retrovirus que una vez integrados en el genoma del huésped han perdido su capacidad de transmisión horizontal (Coffin et al., 1997). Existen antecedentes de la presencia de este tipo de repeticiones en ratón, como es el caso de un tipo de retrovirus endógeno -MaLR no autónomo- que tiene 388.000 copias presentes en el genoma. No se ha podido asignar RVS a ningún cromosoma de ratón, por lo que cabe la posibilidad de que sea un sitio de origen retroviral con numerosas copias repartidas en el genoma. De este modo, aunque el sitio de inserción sea aparentemente el mismo, pTRAP podría estar insertado en RVS situados en distintos lugares del genoma. Si el elemento repetido incluye la secuencia diana de la enzima de digestión empleada en el *southern blot* explicaría que la banda detectada sea la misma en todos los casos de inserción en RVS.

La secuenciación de OrangeSTOP específicamente insertada en RVS confirmó la presencia de mutaciones características de AID. Esto demuestra que este sitio no sólo es un punto caliente de integración de ADN exógeno, sino que también es permisivo con la actividad de AID, lo que parece indicar que los mecanismos que facilitan la integración podrían ser responsables también de favorecer la accesibilidad de AID a estas secuencias. En este sentido, aunque no hay estudios que correlacionen marcas epigenéticas particulares con la preferencia de integración de ADN exógeno, se ha descrito mediante análisis ChIPseq que AID interacciona preferentemente con regiones de ADN asociadas a marcas epigenéticas de accesibilidad

transcripcional, tales como H3K4me3 o H3K27Ac (Yamane et al., 2011) por lo que sería interesante analizar la presencia de estas marcas en RVS.

Diversos estudios han demostrado que la carga de mutaciones introducidas por AID en un transgén correlacionaba con su intensidad de transcripción (Bachl et al., 2001; Fukita et al., 1998; Yoshikawa et al., 2002). En el sistema de monitorización OrangeSTOP desarrollado en este proyecto la carga de mutaciones se estima en función del porcentaje de células Orange⁺ obtenidas y se correlaciona con la intensidad transcripcional en los experimentos de transducción retroviral del casete reportero OrangeSTOP-GFP en células que expresan AID-ER. A pesar de que encontramos este tipo de correlación entre clones transducidos con el vector pMX-PIE, no existe tal correlación en los experimentos de “caza” de promotores. Esto podría sugerir que esta correlación es solamente observable cuando se comparan variaciones en niveles transcripcionales dentro de un mismo contexto de secuencias reguladoras y reforzaría la teoría expuesta anteriormente sobre las múltiples copias de RVS repartidas en el genoma y controladas por distintos promotores.

Especificidad local de secuencia de AID

En este trabajo se ha analizado la especificidad local de secuencia de AID. Numerosos estudios, muchos de ellos anteriores a la identificación de AID, han abordado el patrón de mutaciones asociado al proceso de la SHM en células B, con el fin de encontrar contextos específicos en la secuencia primaria de los genes de inmunoglobulinas que aportaran pistas sobre los mecanismos moleculares y la especificidad de este proceso. Inicialmente, Rogozin *et al.* encontraron que las mutaciones no se distribuían uniformemente en las secuencias variables de inmunoglobulinas, sino que tendían a concentrarse en motivos RGYW (Rogozin and Kolchanov, 1992). Poco después se demostró que la secuencia complementaria reversa WRCY era también un *hotspot* de SHM, lo que sugería que ambas hebras de ADN eran susceptibles a este proceso (Dorner et al., 1998). Más recientemente se observó que estos motivos WRCY/RGYW también están favorecidos en células no-B en las que se inducía la SHM mediante la expresión heteróloga de AID (Yoshikawa et al., 2002). Un reanálisis posterior de estos datos ha propuesto redefinir el *hotspot* primario de SHM como WRCH/DGYW (Rogozin and Diaz, 2004). Por otra parte, AGCT se ha considerado como un *hotspot* primordial de la actividad de AID, al menos en el contexto del CSR (Zarrin et al., 2004). Paralelamente, varios estudios de mutagénesis inducida por AID *in vitro* desvelaron una preferencia intrínseca de esta proteína

por el motivo WRC (Bransteitter et al., 2004; Pham et al., 2003), lo que parece indicar que hay una sutil diferencia entre la especificidad de secuencia de SHM *in vitro* e *in vivo*. Es importante destacar que todas estas observaciones se han hecho a partir de un número de mutaciones limitado, normalmente de pocas decenas. En nuestro caso, utilizamos NGS para realizar un análisis exhaustivo de la distribución de mutaciones en un transgén expresado en células no-B. Las mutaciones inducidas por AID en pares AT en NIH-3T3, al igual que en otras líneas celulares, son muy escasas, por lo que nuestro análisis se restringió a las mutaciones introducidas en citosinas (o guaninas en la hebra contraria) incluídas en *hotspots* WRCY/RGYW. Aunque las causas para este patrón exclusivo de mutaciones en pares GC no se han determinado, parece reflejar la acción conjunta de AID y UNG en ausencia de actividad MMR con tendencia a error. Como se había demostrado anteriormente, las secuencias WRCY/RGYW son en su conjunto motivos en los que AID introduce mutaciones preferencialmente. Sin embargo, nosotros encontramos una gran variabilidad de cada combinación particular respecto a su susceptibilidad a ser mutado por AID. Por ejemplo, la citosina incluida en AGCT se muta con una frecuencia 20 veces superior a la incluida en TACC. De hecho varios de los motivos presentan una frecuencia de mutaciones similar o incluso inferior a la media de mutaciones en citosinas y guaninas encontradas a lo largo de toda la secuencia, lo que indica que algunas combinaciones WRCY/RGYW no se comportan como *hotspots* de SHM en absoluto. Además, al analizar los cuatro *hotspots* más mutados (AGCT, AGCT, AGCA y AGCC) observamos que predomina la A y la G en las posiciones -2 y -1 de la C mutada y la T en la posición +1 (si tenemos en cuenta que en los *hotspots* RGYW el contexto de secuencia de la C mutada se corresponde con el de la hebra contraria), lo que convierte a AGCT en la combinación favorita en ambas hebras. Este resultado demuestra una especificidad de secuencia aún más estricta y establece a AGCT como un *hotspot* favorito para la actividad de AID (Betz et al., 1993; Dorner et al., 1998). La reparación con tendencia a error iniciada por UNG en AGCT ayudaría a definir este *hotspot* como preferido en hipermutación somática, como discutiremos más adelante. Es importante mencionar la similitud encontrada en células B y no-B en la distribución de mutaciones en las secuencias WRCY/RGYW en las que los cuatro *hotspots* más mutados coinciden en ambos casos (AGCT, AGCT, AGCC, AGCA). Esto sugiere la presencia de un mecanismo universal en el *targeting* de AID en pares GC, que es independiente del contexto celular, del gen diana y presumiblemente del nivel de expresión de AID, al menos hasta el alcanzado con nuestro sistema. Como se ha discutido anteriormente, esto sugiere que los cambios en los niveles de AID, a pesar de su impacto cuantitativo en SHM, CSR y la generación de translocaciones cromosómicas (Sernandez et al., 2008), no tienen un efecto cualitativo importante en términos de especificidad de secuencia.

Contribución de UNG a la especificidad local de AID

UNG es una glicosilasa cuya función canónica consiste en reparar los uracilos introducidos en al ADN de forma espontánea, iniciando el proceso de BER (Hagen et al., 2006; Visnes et al., 2009). En este proceso está implicada Pol β , que se encarga de introducir el nucleótido correcto en sustitución del uracilo. Sin embargo, cuando el desapareamiento U:G se genera debido a la desaminación de una citosina por AID, UNG inicia un proceso de reparación con tendencia a error que finaliza con la generación de una mutación en el caso de la SHM, o con una rotura de ADN de doble cadena en el caso de la CSR. En consecuencia, la deficiencia de UNG da lugar a una distorsión en el patrón de mutaciones en SHM y a un bloqueo severo del CSR. Esta actividad mutagénica de UNG sigue siendo una de las grandes preguntas en el campo. En este trabajo hemos descrito que la inhibición de UNG bloquea completamente la generación de transversiones en pares GC, lo que indica que esta glicosilasa es la única que interviene en la generación de este tipo de mutaciones a partir del desapareamiento U:G introducido por AID en células no-B, al igual que en células B (Di Noia et al., 2006; Rada et al., 2004). Por lo tanto, la generación de transversiones en pares GC a partir del uracilo generado por la desaminación inducida por AID parece seguir un mecanismo universal, independiente también del tipo celular. Estos resultados también sugieren que la resolución hacia transversiones de las desaminaciones producidas por AID podría estar vinculada a la presencia de AID en sí misma, presumiblemente debida a una interacción física entre UNG y la desaminasa (Ranjit et al., 2011).

Nuestros experimentos también han revelado que en el contexto de las desaminaciones producidas por AID, la inhibición de UNG provoca un aumento en la frecuencia total de mutaciones. Esto se había sugerido previamente en otros trabajos en los que se inhibía UNG en una línea celular B de pollo (DT40) (Di Noia and Neuberger, 2002) o se analizaba el patrón de mutaciones en ratones deficientes para UNG (Rada et al., 2002), aunque las implicaciones de este resultado no se han analizado en detalle. Nuestros datos demuestran que UNG no sólo desencadena el proceso de reparación con tendencia a error sino que también repara fielmente una parte importante de los desapareamientos U:G generados. Nuestra hipótesis inicial de trabajo fue considerar que los niveles de UNG podrían ser limitantes a la hora de garantizar una correcta reparación de la lesión generada por AID. Sin embargo, los experimentos de sobreexpresión de UNG indican que un aumento en los niveles de la glicosilasa no implica un aumento en la tasa global de reparación fiel. Por el contrario, una mayor actividad de UNG conlleva un aumento en la frecuencia de transversiones y una disminución

en la de transiciones en pares GC, de manera inversa al patrón observado al inhibir UNG. Este resultado parece indicar que al aumentar los niveles de UNG se favorecen las dos vías alternativas de reparación comentadas anteriormente: 1) reparación fiel, que implica la disminución de transiciones en pares GC que normalmente se producirían al replicarse los uracilos generados por la actividad de AID y 2) reparación con tendencia a error, que implica un aumento en las transversiones en pares GC posiblemente asociado al reclutamiento de polimerasas TLS (Fig. 23). Por lo tanto, los niveles de UNG no son limitantes para asegurar una correcta reparación de los desapareamientos U:G generados por AID.

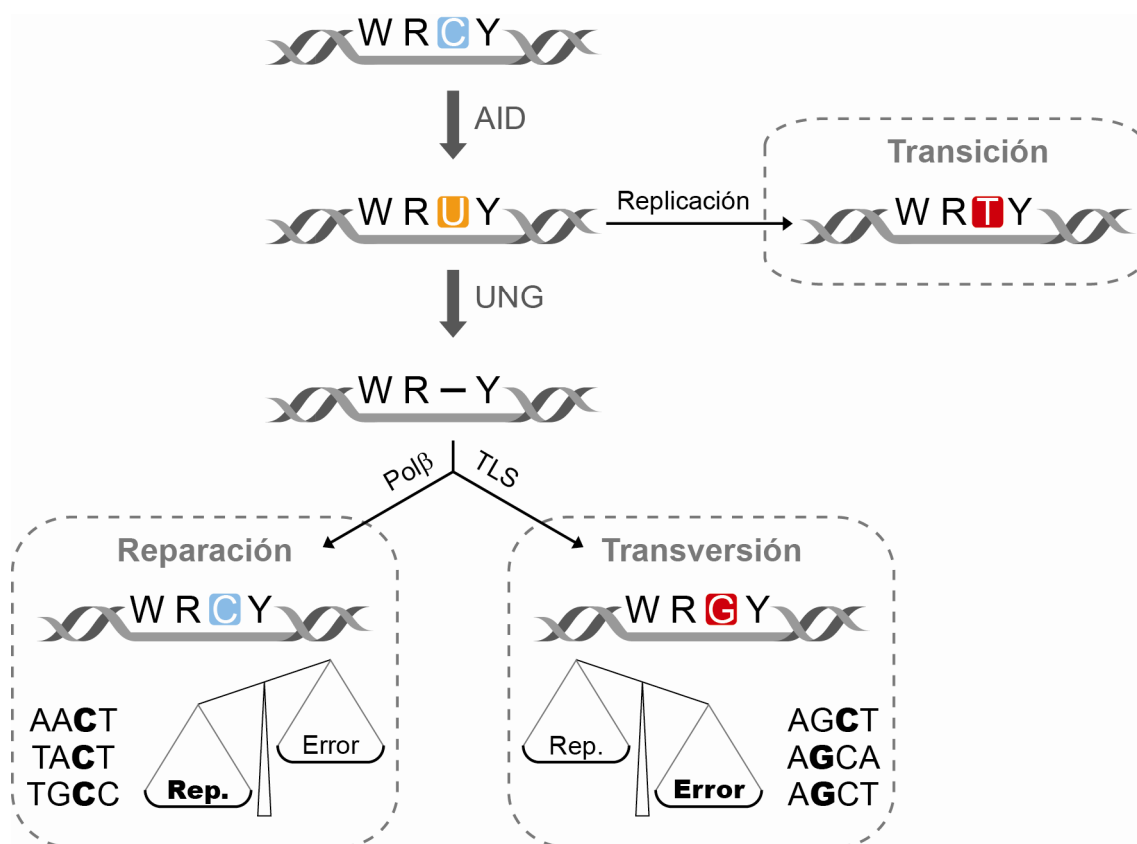


Figura 23: UNG define la especificidad de las mutaciones inducidas por AID

Se representa un esquema en el que se resume los hallazgos principales de la tercera parte de este trabajo (ver texto para más detalles).

Nuestra observación de que UNG puede iniciar ambos procesos nos llevó a preguntarnos si existía alguna especificidad de secuencia a la hora de decantarse por uno de ellos. De hecho, se encontraron importantes diferencias en distintos *hotspots* WRCY/RGYW en cuanto a su susceptibilidad a ser reparados fielmente o con tendencia a error. En concreto, los *hotspots* con una T en la posición +1 de la C mutada presentan una mayor reparación con tendencia a error. Esta es la primera evidencia de la existencia de un mecanismo molecular, aparte de la

propia actividad de AID, que contribuye a la especificidad de la SHM. Es interesante que nuestros resultados podrían explicar la discrepancia entre los *hotspots* de hipermutación definidos *in vivo* e *in vitro*; la tendencia de UNG a desencadenar una reparación mutagénica en la posición +1 de la citosina desaminada conllevaría el sesgo observado hacia secuencias WRCY en contraposición a la preferencia por WRC encontrada *in vitro*. Un trabajo reciente ha demostrado que AID puede mutar el 25% de los genes expresados en células B de centro germinal (Liu et al., 2008). Los autores encontraron que en ausencia de UNG algunos de estos genes aumentaban su tasa de mutación y otros la disminuían, lo que se interpretó como una preferencia de UNG por iniciar una u otra vía de procesamiento –fiel o con tendencia a error– en genes particulares. Sin embargo no se profundizó en los mecanismos que podrían dar cuenta de esta susceptibilidad diferencial. A la vista de nuestros resultados, podemos especular que este comportamiento podría estar determinado, al menos parcialmente, por la secuencia primaria de los distintos genes. En ese sentido, sería interesante analizar el contenido en los diferentes *hotspots* de reparación y mutación de UNG definidos aquí en la colección de secuencias analizadas en el estudio de Liu et al.

Hay que tener en cuenta que la función *per se* de UNG consiste en escindir el uracilo generado, por lo que el hecho de que el contexto de la secuencia primaria influya en las actividades de procesamiento posteriores a la generación del sitio abásico, presumiblemente las de pol β (reparación) o TLS (mutagénica) es en cierto modo sorprendente. Esta contribución de UNG a la resolución del sitio abásico podría explicarse si el contexto de la secuencia entorno al uracilo diana de UNG –en particular en 3'– fuera capaz de alterar el tiempo de residencia o el rango dinámico de UNG sobre el ADN permitiendo su interacción con polimerasas TLS, reclutamiento que favorecería la generación de una transversión. A pesar de que no se ha descrito ninguna interacción entre UNG y polimerasas TLS, existen ejemplos de interacción entre componentes de maquinaria de reparación y este tipo de polimerasas, como es el caso de Msh2-Msh6 y la ADN polimerasa η en la generación de mutaciones AT durante la segunda fase de la SHM (Zeng et al., 2001). Rev1 es una polimerasa TLS cuya actividad asociada a sitios abásicos implica la generación de transversiones, por lo que podría ser una candidata a interaccionar con UNG en los contextos de secuencia comentados anteriormente.

Se ha demostrado que UNG es fundamental para la CSR, puesto que ratones deficientes para la glicosilasa tienen este proceso severamente disminuido (Rada et al., 2002). Nuestros experimentos de inhibición de UNG en células B confirman este resultado, estableciendo a UNG como la única glicosilasa con relevancia funcional en la CSR. Sorprendentemente, la sobreexpresión de UNG en células B activadas también conlleva una reducción en la eficiencia

de la CSR. Como hemos dicho anteriormente, la sobreexpresión de UNG favorece las dos vías de reparación (reparación fiel y reparación con tendencia a error) del sitio abásico generado. Nuestros datos sugieren que en células B activadas UNG podría estar canalizando la resolución del desapareamiento U:G hacia esas dos vías a costa de disminuir una tercera vía independiente que finalizaría con la formación del DSB necesario para la CSR lo que indica que el mecanismo por el que el sitio abásico generado por UNG se resuelve hacia la formación de un DSB es independiente de la glicosilasa.

Por lo tanto, UNG es capaz de dirigir en cierta medida los procesos que ocurren después de la eliminación del sitio abásico de una manera dependiente de la secuencia primaria de ADN, lo que abre una nueva vía de regulación en la resolución de los desapareamientos inducidos por AID. Además, este mecanismo parece ser universal, puesto que hemos encontrado un gran parecido en la especificidad de secuencia en células B y no B. Estos resultados suponen un avance importante en la comprensión de la función de AID en la diversificación de anticuerpos y su contribución a la iniciación de procesos neoplásicos.

Conclusiones

Los resultados obtenidos en esta tesis doctoral permiten extraer las siguientes conclusiones

1. El ensayo de revertientes desarrollado en este trabajo permite monitorizar con alta sensibilidad las mutaciones inducidas por AID mediante la detección de células Orange⁺.
2. Se ha generado un sistema de *screening* basado en una construcción sin promotor (pTRAP) para evaluar a nivel genómico la susceptibilidad a la actividad de AID de *loci* transcripcionalmente activos.
3. La inserción de pTRAP en los clones susceptibles a la actividad de AID ocurre siempre en la misma secuencia de origen retroviral
4. Se ha desarrollado un protocolo de secuenciación masiva que permite detectar mutaciones inducidas por AID.
5. La especificidad de secuencia de la SHM es más restringida que la definida por los *hotspots* de hipermutación WRCY/RGYW y no está limitada a un gen ni tipo celular particular.
6. UNG repara fielmente una parte de los desapareamientos U:G inducidos por AID.
7. La sobreexpresión de UNG aumenta dos vías de resolución alternativas de desapareamientos U:G, una de ellas libre de error, que da lugar a una disminución en la frecuencia de transiciones en pares GC y otra propensa a error, que aumenta la frecuencia de transversiones.
8. La resolución de los desapareamientos U:G a través de una u otra vía de reparación está iniciada por UNG y es, al menos parcialmente, dependiente de la secuencia adyacente a la citosina desaminada. Esta es la primera evidencia de que la actividad de UNG puede redefinir la especificidad de la SHM.

Materiales y métodos

Cultivos celulares

Cultivo de líneas celulares

La línea de células empaquetadoras 293-T y la línea de fibroblastos murinos NIH-3T3 se cultivaron en *Dubbelco's Modified Eagles Medium* (DMEM) con 4.5 g/L de glucosa (GIBCO) suplementado con 10% de suero fetal bovino (FBS), penicilina (50 U/ml) y estreptomicina (50 µg/ml). En adelante, nos referiremos a este medio como DMEM completo.

Cultivo de células B primarias

Las células B primarias se obtuvieron del bazo de ratones *wild type*. Para ello se disgregaron sobre un filtro de nylon de 70 µm en RPMI-1640 (*Sigma-Aldrich*) suplementado con 10% de FBS, penicilina (50 U/ml) y estreptomicina (50 µg/ml). En adelante nos referiremos a éste medio como RPMI completo. La suspensión celular obtenida se centrifugó a 400 g durante 10 minutos. A continuación se trataron las células con un buffer de lisis de eritrocitos (*ACK Lysing Buffer*, *BioWhittaker*) durante 4 minutos a temperatura ambiente, se lavaron con RPMI completo frío y se aislaron las células B mediante depleción inmunomagnética con anti-CD43 (*Miltenyi Biotec*). Las células B se cultivaron a una concentración de $1,2 \times 10^6$ células/ml en medio RPMI completo suplementado con 50 µM de 2-βMercaptoetanol (Gibco), 20mM de Hepes (Gibco), 10 ng/ml de interleuquina 4 (IL4, *PeptoTech*) y 25 µg/ml lipopolisacárido (LPS, *Sigma-Aldrich*). Las células se analizaron a diferentes tiempos de cultivo mediante citometría de flujo. Brevemente, se recogieron 2×10^5 , se lavaron con PBS 1X – 1% FCS, se incubaron con anti-IgG1 biotinilado (*BD biosciences*) y estreptavidina-APC (*BD biosciences*) y se analizaron en un equipo *FACSCanto* (*BD Biosciences*) con el *software FACSDiva* (*BD Biosciences*).

Construcción de vectores

Las mutaciones diseñadas para crear un codón de parada en la proteína fluorescente Orange se introdujeron mediante un procedimiento estándar de mutagénesis dirigida, empleando los oligonucleótidos Stop1-S, Stop1-A y Stop2-S, Stop2-A (Tabla 5). OrangeSTOP se clonó en el

vector *pGEM-T Easy vector* (Promega) y se comprobaron mediante secuenciación con el oligonucleótido universal SP6 las mutaciones introducidas. Se extrajo de este vector mediante digestión con *Bam*HI y se clonó en el vector pMX-PIE.

Las construcciones pQCXI-hΔCD4-AIDER o pQCXI-hΔCD4-EQER se han descrito previamente (Ramiro et al., 2006).

Para crear la construcción pTRAP se amplificó por PCR un fragmento del vector pMX-PIE que contiene la secuencia IRES, GFP y el gen de resistencia a puromicina con los oligonucleótidos TrapIRES-S y TraPuroR-A (Tabla 5). El fragmento obtenido se clonó en el vector pIRES2-EGFP (Clontech) utilizando las dianas de *Ase*I y *Not*I. A continuación se amplificó por PCR OrangeSTOP con los oligonucleótidos TrapOrange-S y TrapOrange-A (Tabla 5) y se clonó en las dianas *Ase*I-*Bam*HI de la construcción anterior.

El cDNA de Ugi nos lo proporcionó el Dr Samuel E. Bennett (*Oregon State University*) y se amplificó por PCR con los oligonucleótidos Ugi-S y Ugi-A (Tabla 5). El producto de PCR se clonó en las dianas *Bam*HI-*Xho*I del vector retroviral pMX-PIE, o en la diana *Eco*RI del vector pBABE-Hyg^R.

El cDNA de la isoforma de localización nuclear de UNG (UNG2) se obtuvo del consorcio I.M.A.G.E (*Integrated Molecular Analysis of Genomes and their Expression*, clon 4009947). Se amplificó por PCR con los oligonucleótidos UNG-S y UNG-A (Tabla 5) y se clonó en el vector *pGEM-T Easy vector*. Se extrajo mediante restricción con el enzima *Eco*RI y se clonó en la misma diana del vector pBABE-HygR o pMX-PIE.

Transfección estable de células NIH-3T3

Se plaquearon 3×10^5 células por placa en dos placas de cultivo de 100 mm de diámetro. Transcurridas 24 horas se recogieron 1×10^6 células, se centrifugaron a 400 g durante 10 minutos, se resuspendieron en 100 µl de la solución MEF1 junto con 5 µg de la construcción pTRAP linearizada y se nucleofectó empleando el programa U-030. Se plaquearon las células en medio DMEM completo precalentado y se incubaron durante 24 horas. Pasado ese tiempo se sustituyó el medio por DMEM completo fresco con 0,4 µg/mL de puromicina y se cultivaron durante tres días. A continuación se seleccionaron células GFP+ por citometría de flujo

preparativa con el equipo *FACSaria TM* (BD biosciences), separando una célula por pocillo en placas de cultivo de 96 pocillos.

Transducción retroviral de células de ratón

Producción de sobrenadante retroviral

Para obtener los sobrenadantes retrovirales se cotransfectaron de forma transitoria las células empaquetadoras 293-T con los distintos vectores usados en cada experimento, y con un vector empaquetador de retrovirus (pCL-ECO), utilizando un protocolo estándar de formación de complejos de fosfato cálcico-ADN. Brevemente, se plaquearon las células 293-T en placas de cultivo de 6 pocillos a una concentración de 5×10^4 células/pocillo. 24 horas después se prepararon 100 µl por cada pocillo de CaCl_2 0,25 M, 2 µg de pCL-ECO y 2 µg de cada vector retroviral empleado y se burbujearon sobre 100 µl de buffer HeBS (0.28 M NaCl_2 , 50 µM Hepes, 1.5 mM Na_2HPO_4). Tras una incubación de 20 minutos a temperatura ambiente se añadió el medio de transfección sobre el cultivo celular. El medio se sustituyó por medio fresco 16 horas después y los sobrenadantes se recogieron después de otras 24 horas.

Transducción retroviral de células NIH-3T3

Las células se sembraron en placas de cultivo de 6 pocillos a una concentración de 3×10^4 células/pocillo. 24 horas después el medio de cultivo se sustituyó por el sobrenadante retroviral con 8 µg/mL de polibreno (*Sigma-Aldrich*). La transducción se realizó durante 2 horas de centrifugación a 1400 g y 16 horas de incubación a 37°C. Pasado ese tiempo se sustituyó el sobrenadante retroviral por medio DMEM fresco con los antibióticos de selección adecuados según el vector utilizado. La eficiencia de transducción se analizó 48 horas después mediante citometría de flujo.

Transducción retroviral de células B primarias

Los cultivos de células B primarias se realizaron como se describió anteriormente y se mantuvieron en condiciones de activación durante 24 horas. Posteriormente el medio de cultivo se sustituyó por el sobrenadante retroviral con 8 µg/mL de polibreno, 25 µg/ml de LPS

y 10 ng/ml de IL4. Las células se centrifugaron durante dos horas a 1400 x g y se incubaron durante 16 horas. Pasado ese tiempo se retiró el sobrenadante retroviral, restituyéndose el medio de activación empleado inicialmente.

Análisis de la expresión de ARN mensajero por PCR cuantitativa a tiempo real (RT-qPCR)

Se extrajo ARN de 1×10^6 células con el reactivo *TRIzol* (Invitrogen), siguiendo las instrucciones del fabricante. A partir de 1 µg de ARN se generó ADN complementario mediante una reacción de transcripción reversa (*Superscript II Reverse Transcriptase*, Invitrogen), utilizando hexámeros al azar como cebadores de la reacción (*Random Hexamers*, Applied Biosystems). El ADN copia (ADNc) obtenido se utilizó para amplificar AID por RT-qPCR, usando *SYBR green PCR Master Mix* (Applied Biosystems) y 0,4 µM de los oligonucleótidos AID-S y AID-A (tabla 5). Se normalizó respecto a la expresión del gen GAPDH en cada muestra. La amplificación se realizó en un equipo *7900HT Fast Real-Time PCR System* (Applied Biosystems)

PCR inversa

Se lisaron 1×10^6 células en buffer Tris-HCl 50 mM, NaCl 200 mM, EDTA 10 mM, SDS 1% y proteinasa K durante 16 horas a 55°C. Se realizó una extracción de ADN con fenol-cloroformo-isoamílico (25:24:1) y se precipitó en etanol usando el coprecipitante *pellet paint* (Novagen), siguiendo las instrucciones del fabricante. Se resuspendió en tampón Tris-HCl 10 mM, pH 8,5.

Se realizaron dos rondas de PCR con los oligonucleótidos iPCR-S, iPCR-A (primera ronda) y iPCR2-S, iPCR2-A (segunda ronda) (Tabla 5) utilizando la polimerasa *Expand Long Template Enzyme* (Roche). Las condiciones de la reacción fueron: Primera ronda; 96°C, 5 minutos; se añade enzima; 80°C, 5 minutos; 25 ciclos de; 94°C, 15 segundos; 59°C, 15 segundos, 68°C, 7 minutos, añadiendo 20 segundos por ciclo. 68°C, 7 minutos de expansión final. Segunda ronda; 96°C, 2 minutos; 25 ciclos de; 94°C, 15 segundos; 58°C, 15 segundos; 68°C, 4 minutos, añadiendo 20 segundos por ciclo; 68°C, 7 minutos de expansión final.

Tabla 5: Secuencias de los oligonucleótidos empleados

Oligonucleótido	Secuencia (5' - 3')
AID-S	ACCTTCGCAACAAGTCTGGCT
AID-A	AGCCTTGCGGTCTTCACAGAA
GAPDH-S	TGAAGCAGGCATCTGAGGG
GAPDH-A	CGAAGGTGGAAGAGTGGGAG
iPCR-S	AGCTGAAGGACGGCGGCCAC
iPCR-A	CCCAGCCCATGGTCTTCTTC
iPCR2-S	GGACTACACCATCGTGGAACA
iPCR2-A	TCACCTTGTAGATGAACTCGC
Orange-S	AATAACATGGCCATCATCAAGGA
Orange1-A	ACGTAGCGGCCCTCGAGTCTCTC
RVS-S	AGAACCTAGAACCTCGCTGG
Orange2-A	CCTTGATGATGGCCATGTTATT
S _μ -S	AATGGATACCTCAGTGGTTTTTAATGGTGGGTTTA
S _μ -A	GCGGCCCGGCTCATTCCAGTTCATTACAG
Stop1-S	AAGGATCCCCACCATGGTGAGCAAGGGCGAGGAGAATAACA
Stop1-A	TGTCGGCGGGGTGCTTCAGCTAGGCCTTGGAGCC
Stop2-S	GGCTCCAAGGCCTAGCTGAAGCACCCCGCCGACA
Stop2-A	AAGGATCCTTACTTGTACAGCTCGTCCATGC
TrapIRES-S	TCATTAATTTATCCAGGATCCCGCTACGTAAATTCCGCCCC
TraPuroR-A	ATGCGGCCGCTCAGGCACCGGGCTTGCGGGTCAT
TrapOrange-S	GCATTAATCCACCATGGTGAGCAAGGGCGAGGAGA
TrapOrange-A	AAGGATCCTTACTTGTACAGCTCGTCCATGC
Ugi-S	GGATCCCGCCACCATGACAAATTTATCTGACATCATTGAAAAAG
Ugi-A	CTCGAGTTATAACATTTTAATTTTATTTTCTCCAT
UNG-S	AAGGATCCCACCATGATCGGCCAGAAGACCCT
UNG-A	AAGGATCCGTACAGCTCCTTCCAGTTGA

Amplificación por PCR de orangeSTOP y secuenciación por Sanger

Se extrajo ADN a partir de 1×10^6 células, como se describe anteriormente. Se realizó una PCR a partir de 1×10^4 células con los oligos Orange-S y Orange1-A (Tabla 5) en un volumen de reacción de 50μl, utilizando la polimerasa Pfu Ultra (*Stratagene*). Las condiciones de la

reacción fueron: 94°C - 5 minutos; 26 ciclos de; 94°C - 10 segundos, 60°C – 30 segundos, 72°C – 1 minuto; 72°C - 7 minutos. El fragmento amplificado se separó en un gel de agarosa 1% y se purificó en columna (*illustra GFX™ Gel Band Purification Kit*), siguiendo las instrucciones del fabricante. A continuación se clonó en el vector *pGEM-T Easy vector* y se secuenciaron clones individuales utilizando el oligonucleótido SP6. Las secuencias obtenidas se analizaron con el programa *Seqman (Lasergene)*.

Para secuenciar orangeSTOP en el sitio retroviral encontrado en la parte de “caza” de promotores activos se siguió el mismo protocolo. Los oligonucleótidos para la reacción fueron RV-S y Orange1-A (Tabla 5).

Para amplificar el sitio de inserción encontrado en la parte de “caza” de promotores se utilizó el enzima *taq DNA polymerase (Roche)*. Se realizó la PCR con los oligonucleótidos RVS-S y Orange2-A (Tabla 5). Las condiciones de la reacción fueron: 94°C - 2 minutos; 30 ciclos de; 94°C - 10 segundos, 60°C – 30 segundos, 68°C – 1 minuto; 68°C - 7 minutos.

Identificación de mutaciones en orangeSTOP por secuenciación masiva

Se amplificó en las mismas condiciones que las descritas anteriormente para secuenciar mediante el método *Sanger*. Se juntaron cinco reacciones distintas por cada muestra (5×10^4 células). Los productos de amplificación se fragmentaron en segmentos de 100-300 bp utilizando un fragmentador *Covaris S2 shearing instrument*. Se añadieron unos adaptadores específicos a cada muestra para su posterior identificación. A continuación se purificaron las distintas muestras utilizando el kit *Agencourt AMPure XP (Beckman Coulter Genomics)*, siguiendo las instrucciones del fabricante. Se amplificaron 10 ciclos más los distintos productos usando oligonucleótidos específicos proporcionados por Illumina, con la polimerasa *Phusion High-Fidelity (Finnzymes)*. Las librerías purificadas se secuenciaron en un *Genome Analyzer IIx (Illumina)*, usando los reactivos *SBS TruSeq v5*, siguiendo las instrucciones del fabricante. Se excluyeron todas las lecturas con una calidad $Q < 20$. El resto de secuencias obtenidas se alinearon utilizando el programa *Novoalign (Novocraft technologies)* y se procesaron con SAMtools (Li et al., 2009). Para obtener una tabla en formato *excel* con todas las mutaciones obtenidas los resultados se *parsearon* con un programa específico diseñado en *Perl*.

Amplificación de S μ e identificación de mutaciones por secuenciación masiva

Se aislaron células CD43-, como se describe anteriormente, de ratones *wild type* y AID *knock-out* y se pusieron en cultivo en condiciones de activación. Se extrajo ADN y se realizó una PCR a partir de $1,3 \times 10^4$ células con los oligonucleótidos S μ -S y S μ -A (Tabla 5) utilizando la polimerasa Pfu Ultra. Las condiciones de reacción fueron 94°C - 5 minutos; 25 ciclos de; 94°C - 30 segundos, 60°C - 30 segundos, 72°C - 50 segundos; 72°C - 7 minutos. Se juntaron 8 reacciones por cada muestra y se procesaron y secuenciaron como se describe para OrangeSTOP.

Southern blot

Se extrajo ADN de los distintos clones, como se describe anteriormente y se digirieron 20 μ g con el enzima de restricción *EcoRV* durante toda la noche. El producto de digestión se separó en un gel de agarosa al 1% a 40 V durante 16 horas. A continuación se realizaron tres lavados del gel con los siguientes buffers, siempre en agitación: HCl 0,25 M durante 15 minutos, buffer desnaturalizante (NaCl 1,5 M, NaOH 0,5 M) durante 45 minutos y buffer neutralizante (NaCl 1,5 M, Tris-HCl 0,5 M pH 7.0) durante 30 minutos. Posteriormente el ADN se transfirió del gel a una membrana de nitrocelulosa (*Amersham Hybondtm-N*, *GE Healthcare*) mediante capilaridad en SSC 10X durante toda la noche. Al día siguiente se fijó covalentemente el ADN a la membrana mediante luz ultravioleta (*UV stratalinker 2400*, *Stratagene*). Se lavó la membrana en SSC 2X durante 10 minutos, y se mantuvo a 65 °C durante una hora en buffer de hibridación (*Amersham Rapid-Hyb buffer*, *GE healthcare*). A continuación se añadió la sonda GFP, un oligonucleótido que hibrida con la proteína fluorescente, previamente marcada radiactivamente utilizando un kit comercial (*Amersham RediprimeTM II Random prime Labelling System*, *GE Healthcare*) y se incubó a 65°C durante toda la noche. Al día siguiente se lavó la membrana tres veces en cada uno de los siguientes buffers: SSC 2X - SDS 0,1% - 15 minutos, SSC1X - SDS 0,1% - 15 minutos. Por último se lavó una vez en SSC 2X durante 5 minutos. Se expuso la membrana en una pantalla *phosphorimager* durante al menos 2 horas, y se reveló utilizando un *Typhoon trio* (*GE Healthcare*).

Ensayos de ganancia y pérdida de función de UNG en células B

Se aislaron células CD43⁻ como se describe anteriormente, y se pusieron en cultivo en condiciones de activación. Se realizó una transducción retroviral con los vectores pmxPIE-UNG, pmxPIE-UGI y el vector vacío como control 24 horas después. A las 48, 72 y 96 horas posteriores al cultivo se marcaron las células con un anticuerpo anti-IgG1 de la forma descrita anteriormente, utilizando DAPI como marcador de viabilidad. La eficiencia de cambio de isotipo se analizó por citometría de flujo en las células GFP⁺.

Bibliografía

Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D., and Brinster, R.L. (1985). The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* 318, 533-538.

Babbage, G., Ottensmeier, C.H., Blaydes, J., Stevenson, F.K., and Sahota, S.S. (2006). Immunoglobulin heavy chain locus events and expression of activation-induced cytidine deaminase in epithelial breast cancer cell lines. *Cancer Res* 66, 3996-4000.

Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M., and Olsson, C. (2001). Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J Immunol* 166, 5051-5057.

Bachl, J., and Olsson, C. (1999). Hypermutation targets a green fluorescent protein-encoding transgene in the presence of immunoglobulin enhancers. *Eur J Immunol* 29, 1383-1389.

Barreto, V., Reina-San-Martin, B., Ramiro, A.R., McBride, K.M., and Nussenzweig, M.C. (2003). C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell* 12, 501-508.

Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J.P., Manis, J.P., and Alt, F.W. (2005). The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* 438, 508-511.

Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., *et al.* (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 144, 353-363.

Begum, N.A., Kinoshita, K., Kakazu, N., Muramatsu, M., Nagaoka, H., Shinkura, R., Biniszkiwicz, D., Boyer, L.A., Jaenisch, R., and Honjo, T. (2004). Uracil DNA glycosylase activity is dispensable for immunoglobulin class switch. *Science* 305, 1160-1163.

Betz, A.G., Neuberger, M.S., and Milstein, C. (1993). Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin V genes. *Immunol Today* 14, 405-411.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463, 1042-1047.

Bransteitter, R., Pham, P., Calabrese, P., and Goodman, M.F. (2004). Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. *J Biol Chem* 279, 51612-51621.

Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci U S A* 100, 4102-4107.

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F.W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422**, 726-730.

Coffin, J.M., Hughes, S.H., and Varmus, H.E. (1997). The Interactions of Retroviruses and their Hosts. *Retroviruses*.

Crouch, E.E., Li, Z., Takizawa, M., Fichtner-Feigl, S., Gourzi, P., Montano, C., Feigenbaum, L., Wilson, P., Janz, S., Papavasiliou, F.N., *et al.* (2007). Regulation of AID expression in the immune response. *J Exp Med* **204**, 1145-1156.

de Yebenes, V.G., Belver, L., Pisano, D.G., Gonzalez, S., Villasante, A., Croce, C., He, L., and Ramiro, A.R. (2008). miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J Exp Med* **205**, 2199-2206.

Dedeoglu, F., Horwitz, B., Chaudhuri, J., Alt, F.W., and Geha, R.S. (2004). Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. *Int Immunol* **16**, 395-404.

Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J.C., and Reynaud, C.A. (2005). Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. *J Exp Med* **201**, 1191-1196.

Di Noia, J., and Neuberger, M.S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43-48.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* **76**, 1-22.

Di Noia, J.M., Rada, C., and Neuberger, M.S. (2006). SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J* **25**, 585-595.

Dickerson, S.K., Market, E., Besmer, E., and Papavasiliou, F.N. (2003). AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* **197**, 1291-1296.

Doi, T., Kinoshita, K., Ikegawa, M., Muramatsu, M., and Honjo, T. (2003). De novo protein synthesis is required for the activation-induced cytidine deaminase function in class-switch recombination. *Proc Natl Acad Sci U S A* **100**, 2634-2638.

Dorner, T., Foster, S.J., Farner, N.L., and Lipsky, P.E. (1998). Somatic hypermutation of human immunoglobulin heavy chain genes: targeting of RGYW motifs on both DNA strands. *Eur J Immunol* **28**, 3384-3396.

Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.H., Robbiani, D.F., Di Virgilio, M., Reina San-Martin, B., Heidkamp, G., Schwickert, T.A., *et al.* (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity* **28**, 630-638.

Endo, Y., Marusawa, H., Kinoshita, K., Morisawa, T., Sakurai, T., Okazaki, I.M., Watashi, K., Shimotohno, K., Honjo, T., and Chiba, T. (2007). Expression of activation-induced

cytidine deaminase in human hepatocytes via NF-kappaB signaling. *Oncogene* 26, 5587-5595.

Fukita, Y., Jacobs, H., and Rajewsky, K. (1998). Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* 9, 105-114.

Gonda, H., Sugai, M., Nambu, Y., Katakai, T., Agata, Y., Mori, K.J., Yokota, Y., and Shimizu, A. (2003). The balance between Pax5 and Id2 activities is the key to AID gene expression. *J Exp Med* 198, 1427-1437.

Gostissa, M., Alt, F.W., and Chiarle, R. (2011). Mechanisms that promote and suppress chromosomal translocations in lymphocytes. *Annu Rev Immunol* 29, 319-350.

Haber, J.E. (1999). DNA repair. Gatekeepers of recombination. *Nature* 398, 665, 667.

Hagen, L., Pena-Diaz, J., Kavli, B., Otterlei, M., Slupphaug, G., and Krokan, H.E. (2006). Genomic uracil and human disease. *Exp Cell Res* 312, 2666-2672.

He, G., and Karin, M. (2011). NF-kappaB and STAT3 - key players in liver inflammation and cancer. *Cell Res* 21, 159-168.

Imai, K., Slupphaug, G., Lee, W.I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H.E., *et al.* (2003). Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* 4, 1023-1028.

Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., and Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci U S A* 101, 1975-1980.

Jansen, J.G., Langerak, P., Tsaalbi-Shtylik, A., van den Berk, P., Jacobs, H., and de Wind, N. (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J Exp Med* 203, 319-323.

Komori, J., Marusawa, H., Machimoto, T., Endo, Y., Kinoshita, K., Kou, T., Haga, H., Ikai, I., Uemoto, S., and Chiba, T. (2008). Activation-induced cytidine deaminase links bile duct inflammation to human cholangiocarcinoma. *Hepatology* 47, 888-896.

Kotani, A., Kakazu, N., Tsuruyama, T., Okazaki, I.M., Muramatsu, M., Kinoshita, K., Nagaoka, H., Yabe, D., and Honjo, T. (2007). Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. *Proc Natl Acad Sci U S A* 104, 1616-1620.

Kou, T., Marusawa, H., Kinoshita, K., Endo, Y., Okazaki, I.M., Ueda, Y., Kodama, Y., Haga, H., Ikai, I., and Chiba, T. (2007). Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. *Int J Cancer* 120, 469-476.

Kovalchuk, A.L., duBois, W., Mushinski, E., McNeil, N.E., Hirt, C., Qi, C.F., Li, Z., Janz, S., Honjo, T., Muramatsu, M., *et al.* (2007). AID-deficient Bcl-xL transgenic mice develop delayed atypical plasma cell tumors with unusual Ig/Myc chromosomal rearrangements. *J Exp Med* 204, 2989-3001.

Kuppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* 5, 251-262.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.

Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliussen, T., *et al.* (2010). Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* 42, 969-972.

Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451, 841-845.

Luedde, T., and Schwabe, R.F. (2011). NF-kappaB in the liver--linking injury, fibrosis and hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* 8, 108-118.

Manis, J.P., Dudley, D., Kaylor, L., and Alt, F.W. (2002). IgH class switch recombination to IgG1 in DNA-PKcs-deficient B cells. *Immunity* 16, 607-617.

Manis, J.P., Gu, Y., Lansford, R., Sonoda, E., Ferrini, R., Davidson, L., Rajewsky, K., and Alt, F.W. (1998). Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *J Exp Med* 187, 2081-2089.

Martomo, S.A., Fu, D., Yang, W.W., Joshi, N.S., and Gearhart, P.J. (2005). Deoxyuridine is generated preferentially in the nontranscribed strand of DNA from cells expressing activation-induced cytidine deaminase. *J Immunol* 174, 7787-7791.

Matsumoto, Y., Marusawa, H., Kinoshita, K., Endo, Y., Kou, T., Morisawa, T., Azuma, T., Okazaki, I.M., Honjo, T., and Chiba, T. (2007). Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med* 13, 470-476.

Maul, R.W., Saribasak, H., Martomo, S.A., McClure, R.L., Yang, W., Vaisman, A., Gramlich, H.S., Schatz, D.G., Woodgate, R., Wilson, D.M., 3rd, *et al.* (2011). Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat Immunol* 12, 70-76.

McBride, K.M., Barreto, V., Ramiro, A.R., Stavropoulos, P., and Nussenzweig, M.C. (2004). Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J Exp Med* 199, 1235-1244.

McBride, K.M., Gazumyan, A., Woo, E.M., Barreto, V.M., Robbiani, D.F., Chait, B.T., and Nussenzweig, M.C. (2006). Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proc Natl Acad Sci U S A* 103, 8798-8803.

Merrihew, R.V., Marburger, K., Pennington, S.L., Roth, D.B., and Wilson, J.H. (1996). High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome. *Mol Cell Biol* 16, 10-18.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem* 279, 52353-52360.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102, 553-563.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* 274, 18470-18476.

Ngo, V.N., Young, R.M., Schmitz, R., Jhavar, S., Xiao, W., Lim, K.H., Kohlhammer, H., Xu, W., Yang, Y., Zhao, H., *et al.* (2011). Oncogenically active MYD88 mutations in human lymphoma. *Nature* 470, 115-119.

Okazaki, I.M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. *J Exp Med* 197, 1173-1181.

Orthwein, A., Patenaude, A.M., Affar el, B., Lamarre, A., Young, J.C., and Di Noia, J.M. (2010). Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. *J Exp Med* 207, 2751-2765.

Pan-Hammarstrom, Q., Jones, A.M., Lahdesmaki, A., Zhou, W., Gatti, R.A., Hammarstrom, L., Gennery, A.R., and Ehrenstein, M.R. (2005). Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells. *J Exp Med* 201, 189-194.

Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., Muramatsu, M., Honjo, T., Morse, H.C., 3rd, Nussenzweig, M.C., and Dalla-Favera, R. (2008). AID is required for germinal center-derived lymphomagenesis. *Nat Genet* 40, 108-112.

Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006). PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc Natl Acad Sci U S A* 103, 395-400.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Kuppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412, 341-346.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., *et al.* (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 143, 122-133.

Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* 418, 99-103.

Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 424, 103-107.

Phung, Q.H., Winter, D.B., Cranston, A., Tarone, R.E., Bohr, V.A., Fishel, R., and Gearhart, P.J. (1998). Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J Exp Med* 187, 1745-1751.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463, 1101-1105.

Porta, C., Riboldi, E., and Sica, A. (2011). Mechanisms linking pathogens-associated inflammation and cancer. *Cancer Lett* 305, 250-262.

Prakash, S., Johnson, R.E., and Prakash, L. (2005). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem* 74, 317-353.

Puente, X.S., Pinyol, M., Quesada, V., Conde, L., Ordonez, G.R., Villamor, N., Escaramis, G., Jares, P., Bea, S., Gonzalez-Diaz, M., *et al.* (2011). Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 475, 101-105.

Rada, C., Di Noia, J.M., and Neuberger, M.S. (2004). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* 16, 163-171.

Rada, C., Ehrenstein, M.R., Neuberger, M.S., and Milstein, C. (1998). Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* 9, 135-141.

Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* 12, 1748-1755.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135, 1201-1212.

Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature* 381, 751-758.

Ramiro, A., Reina San-Martin, B., McBride, K., Jankovic, M., Barreto, V., Nussenzweig, A., and Nussenzweig, M.C. (2007). The role of activation-induced deaminase in antibody diversification and chromosome translocations. *Adv Immunol* 94, 75-107.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W., *et al.* (2006). Role of genomic instability and p53 in AID-induced c-myc-IgH translocations. *Nature* 440, 105-109.

Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A., and Nussenzweig, M.C. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* 118, 431-438.

Ramiro, A.R., Stavropoulos, P., Jankovic, M., and Nussenzweig, M.C. (2003). Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat Immunol* 4, 452-456.

Ranjit, S., Khair, L., Linehan, E.K., Ucher, A.J., Chakrabarti, M., Schrader, C.E., and Stavnezer, J. (2011). AID binds cooperatively with UNG and Msh2-Msh6 to Ig switch regions dependent upon the AID C terminus. *J Immunol* 187, 2464-2475.

Rassool, F.V., McKeithan, T.W., Neilly, M.E., van Melle, E., Espinosa, R., 3rd, and Le Beau, M.M. (1991). Preferential integration of marker DNA into the chromosomal fragile site at 3p14: an approach to cloning fragile sites. *Proc Natl Acad Sci U S A* 88, 6657-6661.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., *et al.* (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 102, 565-575.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A., *et al.* (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* 135, 1028-1038.

Rogozin, I.B., and Diaz, M. (2004). Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J Immunol* 172, 3382-3384.

- Rogozin, I.B., and Kolchanov, N.A. (1992). Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim Biophys Acta* 1171, 11-18.
- Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N.J., Hentschke, M., Chen, W., Pu, F., Peng, Y., Li, J., *et al.* (2011). Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N Engl J Med* 365, 718-724.
- Ross, A.L., and Sale, J.E. (2006). The catalytic activity of REV1 is employed during immunoglobulin gene diversification in DT40. *Mol Immunol* 43, 1587-1594.
- Sayegh, C.E., Quong, M.W., Agata, Y., and Murre, C. (2003). E-proteins directly regulate expression of activation-induced deaminase in mature B cells. *Nat Immunol* 4, 586-593.
- Schatz, D.G. (2004). V(D)J recombination. *Immunol Rev* 200, 5-11.
- Schrader, C.E., Linehan, E.K., Mochegova, S.N., Woodland, R.T., and Stavnezer, J. (2005). Inducible DNA breaks in Ig S regions are dependent on AID and UNG. *J Exp Med* 202, 561-568.
- Schreck, S., Buettner, M., Kremmer, E., Bogdan, M., Herbst, H., and Niedobitek, G. (2006). Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. *J Pathol* 210, 26-31.
- Sernandez, I.V., de Yebenes, V.G., Dorsett, Y., and Ramiro, A.R. (2008). Haploinsufficiency of activation-induced deaminase for antibody diversification and chromosome translocations both in vitro and in vivo. *PLoS One* 3, e3927.
- Shaffer, A.L., Lin, K.I., Kuo, T.C., Yu, X., Hurt, E.M., Rosenwald, A., Giltnane, J.M., Yang, L., Zhao, H., Calame, K., *et al.* (2002). Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* 17, 51-62.
- Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol* 22, 1567-1572.
- Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* 280, 1750-1752.
- Soulas-Sprauel, P., Le Guyader, G., Rivera-Munoz, P., Abramowski, V., Olivier-Martin, C., Goujet-Zalc, C., Charneau, P., and de Villartay, J.P. (2007). Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination. *J Exp Med* 204, 1717-1727.
- Stavnezer, J. (1996). Antibody class switching. *Adv Immunol* 61, 79-146.
- Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 26, 261-292.

Suzuki, K., Ohbayashi, F., Nikaido, I., Okuda, A., Takaki, H., Okazaki, Y., and Mitani, K. (2010). Integration of exogenous DNA into mouse embryonic stem cell chromosomes shows preference into genes and frequent modification at junctions. *Chromosome Res* 18, 191-201.

Suzuki, S., Ono, R., Narita, T., Pask, A.J., Shaw, G., Wang, C., Kohda, T., Alsop, A.E., Marshall Graves, J.A., Kohara, Y., *et al.* (2007). Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genet* 3, e55.

Takai, A., Marusawa, H., Minaki, Y., Watanabe, T., Nakase, H., Kinoshita, K., Tsujimoto, G., and Chiba, T. (2011). Targeting activation-induced cytidine deaminase prevents colon cancer development despite persistent colonic inflammation. *Oncogene*.

Takizawa, M., Tolarova, H., Li, Z., Dubois, W., Lim, S., Callen, E., Franco, S., Mosaico, M., Feigenbaum, L., Alt, F.W., *et al.* (2008). AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *J Exp Med* 205, 1949-1957.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* 28, 621-629.

Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2010). B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. *Nat Immunol* 11, 148-154.

Van Dyck, E., Stasiak, A.Z., Stasiak, A., and West, S.C. (1999). Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* 398, 728-731.

Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C.K., Stephens, P., Davies, H., Jones, D., Lin, M.L., Teague, J., *et al.* (2011). Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* 469, 539-542.

Visnes, T., Doseth, B., Pettersen, H.S., Hagen, L., Sousa, M.M., Akbari, M., Otterlei, M., Kavli, B., Slupphaug, G., and Krokan, H.E. (2009). Uracil in DNA and its processing by different DNA glycosylases. *Philos Trans R Soc Lond B Biol Sci* 364, 563-568.

Vuong, B.Q., Lee, M., Kabir, S., Irimia, C., Macchiarulo, S., McKnight, G.S., and Chaudhuri, J. (2009). Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nat Immunol* 10, 420-426.

Wang, K., Kan, J., Yuen, S.T., Shi, S.T., Chu, K.M., Law, S., Chan, T.L., Kan, Z., Chan, A.S., Tsui, W.Y., *et al.* (2011). Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet*.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520-562.

Weill, J.C., and Reynaud, C.A. (2008). DNA polymerases in adaptive immunity. *Nat Rev Immunol* 8, 302-312.

Wiesendanger, M., Kneitz, B., Edelmann, W., and Scharff, M.D. (2000). Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J Exp Med* 191, 579-584.

Wilson, T.M., Vaisman, A., Martomo, S.A., Sullivan, P., Lan, L., Hanaoka, F., Yasui, A., Woodgate, R., and Gearhart, P.J. (2005). MSH2-MSH6 stimulates DNA polymerase η , suggesting a role for A:T mutations in antibody genes. *J Exp Med* 201, 637-645.

Wurtele, H., Little, K.C., and Chartrand, P. (2003). Illegitimate DNA integration in mammalian cells. *Gene Ther* 10, 1791-1799.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol* 12, 62-69.

Yang, S.Y., Fugmann, S.D., Gramlich, H.S., and Schatz, D.G. (2007). Activation-induced cytidine deaminase-mediated sequence diversification is transiently targeted to newly integrated DNA substrates. *J Biol Chem* 282, 25308-25313.

Yoshikawa, K., Okazaki, I.M., Eto, T., Kinoshita, K., Muramatsu, M., Nagaoka, H., and Honjo, T. (2002). AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* 296, 2033-2036.

Yu, K., Roy, D., Bayramyan, M., Haworth, I.S., and Lieber, M.R. (2005). Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops. *Mol Cell Biol* 25, 1730-1736.

Zarrin, A.A., Alt, F.W., Chaudhuri, J., Stokes, N., Kaushal, D., Du Pasquier, L., and Tian, M. (2004). An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nat Immunol* 5, 1275-1281.

Zeng, X., Winter, D.B., Kasmer, C., Kraemer, K.H., Lehmann, A.R., and Gearhart, P.J. (2001). DNA polymerase η is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol* 2, 537-541.

Zhou, C., Saxon, A., and Zhang, K. (2003). Human activation-induced cytidine deaminase is induced by IL-4 and negatively regulated by CD45: implication of CD45 as a Janus kinase phosphatase in antibody diversification. *J Immunol* 170, 1887-1893.

Anexos

Publicaciones

Pérez-Durán, P., Belver, L., de Yébenes, V.G., Delgado, P., Pisano, D.G., and Ramiro, A.R. (2011). UNG shapes the specificity of AID-induced somatic hypermutation. J Exp Med, *under review*.

Pérez-Durán, P., de Yébenes, V.G., and Ramiro, A.R. (2007). Oncogenic events triggered by AID, the adverse effect of antibody diversification. Carcinogenesis 28, 2427-2433.

UNG shapes the specificity of AID-induced somatic hypermutation

Pablo Pérez-Durán¹, Laura Belver¹, Virginia G de Yébenes¹, Pilar Delgado¹, David G Pisano² and Almudena R Ramiro^{1*}

¹DNA Hypermutation and Cancer Group and ²Bioinformatics Unit, Spanish National Cancer Research Center (CNIO), Spain.

*Corresponding author:

Almudena R Ramiro

Spanish National Cancer Research Center (CNIO)

Melchor Fernandez Almagro 3, 28029 Madrid, Spain

Ph No 34917328000-3510

Fax No 34917328033

aramiro@cnio.es

Address from December 1 2011:

Centro Nacional de Investigaciones Cardiovasculares (CNIC). aramiro@cnic.es

Short title: UNG mutational specificity

Character count: 30762

Abbreviations used: AID, Activation Induced Deaminase; BER, base excision repair; CSR, class switch recombination; ER, estrogen receptor; NGS, next generation sequencing; OHT, tamoxifen; SHM, somatic hypermutation; S μ , μ switch region; TLS, translesion synthesis; Ugi, uracil-DNA glycosylase inhibitor; UNG, Uracil-N-Glycosylase.

Abstract

Secondary diversification of antibodies is a critical component of the immune response achieved through the somatic hypermutation (SHM) and class switch recombination (CSR) reactions. Activation Induced Deaminase (AID) initiates both processes by deaminating cytosine residues on immunoglobulin genes. The U:G mismatch thus generated can be processed by alternative pathways to give rise to a mutation (SHM) or a DNA double strand break (CSR). Central to this processing is the activity of Uracil-N-Glycosylase (UNG), an enzyme normally involved in error-free base excision repair. Here we have addressed AID specificity making use of a reporter assay and next generation sequence analysis. Loss- and gain- of function experiments show that UNG activity can promote both error-prone and high fidelity repair of U:G initiated lesions. Unexpectedly, the balance between these alternative outcomes is influenced by the sequence context of the deaminated cytosine, with individual hotspots exhibiting higher susceptibility to UNG-triggered error-prone resolution and mutation fixation. Our work reveals UNG as a new molecular layer that shapes the specificity of AID-induced mutations in a sequence dependent fashion. These results have important implications for the understanding of the secondary diversification of antibodies and will provide new insights into the role of AID in cancer development.

Introduction

Activation Induced Deaminase (AID) is a critical component of the immune response, as it is essential for the reactions of antibody secondary diversification, namely somatic hypermutation (SHM) and class switch recombination (CSR) (Muramatsu et al., 2000; Revy et al., 2000), which take place in B cells that have been activated by antigen. SHM introduces nucleotide substitutions in the variable, antigen binding region of immunoglobulins, allowing the generation of variants with higher affinity for antigen that are selected by affinity maturation (reviewed in (Di Noia and Neuberger, 2007; Peled et al., 2008)). CSR is a region specific recombination reaction that replaces the primary μ constant region by a downstream constant region, providing the antibody with specialized means of antigen removal (reviewed in (Stavnezer et al., 2008)). AID deficiency prevents SHM and CSR and promotes a hyper-immunoglobulin M immunodeficiency in humans (Muramatsu et al., 2000; Revy et al., 2000).

AID initiates SHM and CSR by deaminating cytosine residues at the DNA of immunoglobulin genes (Di Noia and Neuberger, 2007; Maul et al., 2011; Petersen-Mahrt et al., 2002). The U:G mismatch thus generated can be processed by alternative pathways to give rise to the introduction of a mutation (SHM) or the generation of a DNA double strand break and a recombination reaction (CSR). If the U:G mismatch is simply replicated over, it will give rise to transition mutations (C>T or G>A, depending on the DNA strand where the deamination took place). Alternatively, uracil can be excised from the DNA by uracil-N-glycosylase (UNG), an enzyme that is part of the base excision repair (BER) pathway. UNG inhibition in B cell lines and UNG deficiency in mice and humans give rise to an altered SHM pattern where the proportion of transition mutations is dramatically increased and transversion mutations (C>A or G and G>C or T) are decreased (Di Noia and Neuberger, 2002; Imai et al., 2003; Rada et al., 2004; Rada et al., 2002; Xue et al., 2006). In addition, UNG deficiency severely impairs CSR (Imai et

al., 2003; Rada et al., 2002). These results showed that in the context of antibody diversification, UNG contributes to the mutagenic resolution of U:G mismatches rather than to their error-free repair by BER, but the basis of this error-prone activity of UNG is not understood. Alternatively, the U:G mispair can be recognized by the Msh2/Msh6 components of the mismatch repair machinery, which can trigger a patch DNA synthesis leading to nucleotide substitutions mostly at A or T nucleotides (also known as second phase of SHM) (Rada et al., 2004; Rada et al., 1998; Schrader et al., 1999; Wiesendanger et al., 2000).

Somatic mutations are not randomly distributed along Ig variable genes; instead, they accumulate preferentially at mutational hotspots. Analysis of Ig gene databases from hypermutating B cells revealed that SHM at cytosine residues is strongly favored when they are part of an RGYW (where R=A/G, Y=C/T, W=A/T) or its reverse complement sequence WRCY (Dorner et al., 1998; Rogozin and Kolchanov, 1992). Among the RGYW/WRCY variants, AGCT has been proposed to be a primordial motif for CSR (Zarrin et al., 2004). A similar hotspot preference has been shown in non-Ig transgenes expressed in B cell lines (Martin and Scharff, 2002) and in fibroblasts with heterologous AID expression (McBride et al., 2004; Yoshikawa et al., 2002), suggesting that targeting of these motifs is a universal feature of the SHM machinery. In addition, biochemical studies showed that AID activity on DNA substrates *in vitro* displays a mutational preference for the WRC motif (Bransteitter et al., 2004; Pham et al., 2003; Yu et al., 2004). Thus, the targeting process of SHM *in vivo* displays sensitivity for the nucleotide that flanks the C at 3' that is not fully recapitulated in the *in vitro* assays.

AID activity is not exclusively restricted to immunoglobulin genes, and can cause relatively widespread DNA lesions in the genome. Mutations bearing the hallmarks of SHM were detected on in the *Bcl6* gene in human memory B cells (Shen et al., 1998) and in a number of

loci –including the proto-oncogenes *PIM1*, *Pax5* and *Myc* in diffuse B cell lymphomas (Pasqualucci et al., 2001). More recently, an extensive sequencing study estimated that 25% of the genes expressed by germinal center B cells can accumulate AID-mediated mutations (Liu et al., 2008). A direct link between AID function and B cell neoplasias was established by the finding that lymphomagenic c-myc/IgH translocations (Adams et al., 1985) are initiated by AID, both *in vivo* and *in vitro* (Dorsett et al., 2007; Ramiro et al., 2006; Ramiro et al., 2004; Robbiani et al., 2008). Indeed, the absence of AID in a number of lymphoma mouse models delays the onset or shifts the nature of the neoplasia (Kovalchuk et al., 2007; Pasqualucci et al., 2008; Ramiro et al., 2004). In addition, there are evidences that AID activity may not be confined to the B cell lineage and could contribute to non-B cell neoplasias. AID expression has been detected in a number of human cancers, which correlated with accumulation of mutations in various genes, including *Trp53* or *Myc* (Endo et al., 2007; Kou et al., 2007; Matsumoto et al., 2007). Recently, it has been shown that AID deficiency exerts protection against the development of colitis-associated cancers (Takai et al., 2011). Therefore AID specificity is a relevant issue for the understanding of both secondary diversification of antibodies and the role of this enzyme in cancer.

Here, we have addressed the contribution of UNG to the specificity of AID-induced mutations by combining gain- and loss- of-function approaches and mutation analysis using next generation sequencing technology. We find that UNG can process U:G lesions generated by AID to give rise both to faithful and error-prone repair depending of the sequence context. Our results provide the first evidence that UNG activity shapes the sequence specificity of AID during SHM.

Results

Assay to monitor AID mutational activity

In order to monitor AID mutational activity we developed a fluorescence revertance assay. Briefly, a stop codon overlapping with an AGCT AID mutational hotspot was introduced at positions 230-233 of the sequence encoding the mOrange fluorescent protein (mOrange^{STOP}, Figure 1a), an mRFP1 variant that can be easily detected by flow cytometry (Shaner et al., 2004). This TAG stop codon generates a non-fluorescent truncated protein, but transversion mutations at its third nucleotide revert it to TAC or TAT tyrosine-encoding codons that reconstitute the full-length mOrange fluorescent protein. mOrange^{STOP} was introduced into the GFP-containing retroviral vector pMX-PIE (Barreto et al., 2003) to allow the tracking of transduced cells. Inducible AID activity was achieved by fusing AID to the estrogen-binding domain of estrogen receptor (AID-ER), thus generating a protein that can be translocated into the nucleus -and therefore grant access to its DNA substrate- by tamoxifen (OHT) treatment (Doi et al., 2003). AID-ER -or the catalitically inactive mutant AID^{E58Q}-ER- were cloned into a second retroviral vector that contains a truncated, signaling-devoid form of the human CD4 molecule (Δ huCD4) (Figure 1a). To test the mOrange^{STOP} revertance assay we retrovirally transduced the mOrange^{STOP} vector along with either AID-ER or AID^{E58Q}-ER containing vectors into NIH-3T3 mouse fibroblasts. After 3 days of puromycin selection > 95% of cells were GFP⁺ Δ huCD4⁺ (not shown). Cells were then cultured in the presence or absence of OHT for up to 11 days. We detected the appearance of mOrange⁺ cells in AID-ER transduced cultures as soon as 2 days after OHT treatment and their percentage increased with time (Figure 1b, c). In contrast, AID^{E58Q}-ER transduction failed to generate detectable mOrange⁺ cells, and in the absence of OHT AID-ER only promoted marginal numbers of mOrange revertants. These results show that AID mutational activity can be monitored by the generation of mOrange revertants in NIH-3T3 cells.

UNG repairs faithfully a fraction of the mutations induced by AID in non-B cells

UNG plays a crucial role in the processing of AID-induced lesions in B cells. In particular, UNG ablation biases the SHM pattern at G/C base-pairs towards transition mutations and impairs CSR (Di Noia and Neuberger, 2002; Rada et al., 2002). To address the role of base excision repair in the resolution of AID-induced lesions in non-B cells we made use of the uracil-DNA glycosylase inhibitor (Ugi) protein from the bacteriophage PSB2 (Di Noia and Neuberger, 2002; Wang and Mosbaugh, 1988). We first tested UNG inhibition by Ugi by transducing spleen mouse B cells with Ugi and measuring CSR after stimulation with LPS and IL4. In agreement with previous findings in UNG deficient mice (Rada et al., 2002), we found that Ugi severely impaired CSR in primary B cells *in vitro*, as revealed by the percentage of IgG1⁺ (Figure 2a, b). Next, we examined the contribution of UNG to the mutational activity of AID in the mOrange-based revertance assay. We co-transduced Ugi together with mOrange^{STOP} and AID-ER into NIH-3T3 cells, induced AID activity by OHT treatment and analyzed the appearance of mOrange⁺ revertant cells for up to 11 days of culture. We found that Ugi expression prevented the generation of mOrange⁺ cells, suggesting that UNG activity is needed to promote the transversion mutations that revert the STOP codon at the mOrange^{STOP} sequence (Figure 2c, d).

To explore the contribution of UNG to the mutational pattern induced by AID in non-B cells, we analysed mutations at the entire mOrange sequence in the presence or absence of Ugi. NIH-3T3 cells were co-transduced with AID-ER, mOrange^{STOP} and either Ugi or control retroviral vectors, cultured in the presence of OHT, and the mOrange^{STOP} transgene was amplified, cloned and sequenced by standard Sanger method. We found that Ugi expression dramatically reduced the proportion of transversions at G/C pairs (control 23% vs Ugi 0.8%, p (Fisher exact test)= 3.5×10^{-7}), and increased the proportion of G/C transitions (control 71% vs Ugi 98%,

$p=2.8 \times 10^{-8}$) (Figure 3a, b), indicating that UNG is required for the generation of G/C transversions in non-B cells, in agreement with the data on immunoglobulin gene mutations in UNG deficient or inhibited B cells (Di Noia and Neuberger, 2002; Rada et al., 2002). Although the proportion of mutations at A/T pairs was reduced in Ugi-expressing cells (Figure 3b), this difference was not statistically significant. Interestingly, we found that UNG inhibition resulted in a significant increase in the overall mutation load (control, 10.0×10^{-4} mutations/bp vs Ugi, 22.9×10^{-4} mutations/bp; p (unpaired t test)=0.014), which is accounted for by an increase in the frequency of transitions at G/C pairs (control, 7.1×10^{-4} mutations/bp vs Ugi 22.5×10^{-4} mutations/bp; p (unpaired t test)=0.0026). We conclude that UNG repairs faithfully a significant fraction of the U:G lesions. Therefore, these results show that UNG can trigger two alternative resolutions of the deamination events initiated by AID: it promotes transversions at G/C pairs and it faithfully repairs a fraction of the lesions.

Detection of AID-induced mutations by next generation sequencing

The finding that Ugi expression increases the mutation load in the mOrange^{STOP} transgene, opened the question of whether enforcing UNG expression can lead to a more efficient repair of U:G lesions initiated by AID and whether this repair activity displays any sequence preference. To address this issue, we set to develop a next generation sequencing (NGS) approach that would allow mutational analysis at very high depth coverage. We PCR-amplified the mOrange^{STOP} transgene from NIH-3T3 cells that had been co-transduced with mOrange^{STOP} and AID-ER or AID^{E58Q}-ER and cultured for 11 days in the presence of OHT. PCR products were sheared, bar-coded and sequenced in an Illumina platform. After high quality filtering and alignment we obtained a depth of 8.7×10^4 reads per base position on average, which provided the detection of roughly 100 mutations per cytosine residue per experiment (Table 1). First, we compared the mutation pattern of the sequences obtained by Sanger and NGS. As expected,

the mutation frequency observed by conventional sequencing was increased at cytosines contained in WRC and $WRCY$ AID mutational hotspots (Figure 4a). The same mutation preference was found in the sequences obtained by NGS, with roughly 4-fold and 6-fold increase in mutation frequency at WRC/GYW and $WRCY/RGYW$ hotspots, respectively (Figure 4b). In contrast, sequences obtained from AID^{E58Q} -ER transduced cells did not show accumulation of mutations at these hotspots and instead harbored evenly distributed mutations (Figure 4b). In successive experiments, analyses will always show the mutation frequencies after subtracting AID^{E58Q} -ER mutations on a per-base basis. These results validate the use of NGS for the detection of AID-mutations at a depth range 100-1000 fold higher than that of conventional sequencing.

We next analyzed the mutation distribution at cytosines and guanines contained in each of the $WRCY$ and $RGYW$ hotspots for SHM *in vivo* in the $mOrange^{STOP}$ sequence (Figure 4c). All hotspot sequences are shown on the coding strand, and therefore in the case of $RGYW$ motifs, the context of the mutated cytosine is that of the reverse complement sequence. We found that the mutation susceptibility differed widely across the different $WRCY$ and $RGYW$ combinations, with the highest mutation frequency accumulating at cytosines at the AGCT motif and guanines at AGCA and AGCT motifs. To approach whether this uneven mutation distribution mimicked the SHM taking place in B cells, we stimulated spleen primary B cells with LPS and IL4 and analyzed mutations at the μ switch (S_μ) region of the immunoglobulin heavy chain locus by NGS. Although the average mutation frequency was sensibly lower in the S_μ region than in the $mOrange^{STOP}$ sequence, we found a remarkably similar preference for individual $WRCY/RGYW$ hotspots (Figure 4d). In particular, $AGCT$, $AGCC$ and $AGCA$ motifs were highly mutated while others, such as $TGCC$, $AGTT$ or $GGCA$ show only average mutation frequencies. We conclude that the mutability of different motifs conforming to the canonical $WRCY/RGYW$

mutational hotspots is highly variable and that this mutation susceptibility is not specific of a particular target gene or cell type.

UNG shapes the specificity of AID-induced mutations

To approach the contribution of UNG activity in the processing of AID-mediated deaminations, we performed lost- and gain- of function experiments. NIH-3T3 cells were co-transduced with AID-ER, mOrange^{STOP} and either Ugi, UNG or control vectors and cultured in the presence of OHT. The effect of UNG overexpression was first assessed with the mOrange^{STOP} reversion assay. In contrast to the dramatic blockade in the generation of mOrange⁺ cells after Ugi expression (Figure 2c), we found that cells transduced with UNG generated a higher percentage of mOrange⁺ revertants (Figure 5a, b). Next, we analyzed mutations at the mOrange^{STOP} transgene by NGS as described above in cells transduced with control, Ugi or UNG vectors. In agreement with the results obtained with conventional sequencing (Figure 3), expression of Ugi resulted in a dramatic increase of transitions and in the almost complete absence of transversions at G/C pairs (Figure 5c, d). In contrast, UNG overexpression promoted a higher proportion of transversions at G/C nucleotides at the expense of G/C transitions (Figure 5c). However, the overall mutation frequency in UNG overexpressing cells was similar to that observed in control cells (Figure 5d). These results show that UNG overexpression does not significantly alter the mutation load, but it does contribute to the final outcome of the resolution of the lesions induced by AID and increases the generation of transversions at G/C pairs.

Our observation that the frequency of mOrange⁺ revertants is increased in UNG overexpressing cells without a substantial increase in overall mutation frequency suggested that the contribution of UNG to the generation of mutations might display some sequence

preference. Indeed, we found that the frequency of transversion mutations at the G²³¹ contained in the TAG stop codon of mOrange^{STOP} was significantly higher in cells that had been transduced with UNG, as compared with control cells (p (paired t test)=0.032) (Figure 6a). As expected, Ugi expression abolished transversion mutations at this position. To determine the distribution of mutations in the mOrange^{STOP} transgene we analyzed the frequency of transition and transversion mutations at cytosine and guanine residues contained in WRCY/RGYW hotspots from control, Ugi and UNG expressing cells. Transition and transversion mutations were analyzed separately for the sake of clarity (Figure 6b). As expected, we found an uneven distribution of mutations across different hotspots. In agreement with the global analysis shown in Figure 5d, Ugi expression resulted in a larger frequency of transition mutations and UNG overexpression generally increased transversion frequency. However, not all hotspots seemed equally sensitive to variations in UNG activity. To quantify this differential sensitivity, we calculated the Ugi/control ratio of transitions and the UNG/control ratio of transversions. We reasoned that, given that UNG inhibition by Ugi increases the frequency of transition mutations, the ratio of transition frequency between Ugi and control cells would reflect the preferred sites for UNG-mediated faithful repair (Figure 6c, upper graph). Conversely, the ratio of transversion mutations between UNG and control cells would indicate preferred hotspots for error-prone repair by UNG (Figure 6c, lower graph). Consistent with this idea, we observed that indeed the Ugi/control transition ratios mirror the UNG/control transversion ratios and that individual types of WRCY/RGYW hotspots display different propensity to faithful *versus* error-prone repair (Figure 6c). We found that two of the most strongly mutated hotspots, cytosines and guanines at AGCT and guanines at AGCA, have a prominent trend to accumulate transversion mutations and are poorly repaired by UNG. In contrast, deaminated cytosines at AACT, TACT and TGCC hotspots seem particularly predisposed to be processed by UNG-initiated error-free repair (Figure 7). We conclude that

UNG activity displays sequence preference that biases the error-free *versus* error-prone resolution of AID initiated U:G mismatches.

Discussion

We have addressed in this study the role of UNG activity in the resolution of the U:G mismatches generated by AID-induced cytosine deamination. To that aim, we implemented a fluorescence reversion assay to monitor AID activity. In contrast to other monitoring systems published previously (Bachl and Olsson, 1999; Yoshikawa et al., 2002), the mOrange^{STOP} reversion assay includes a second fluorescent protein (GFP) that allows tracking of transduced cells, providing a more accurate and sensitive detection of AID-induced revertants. This assay will be particularly useful in primary cells where transduction efficiency is typically low. In addition, we have made use of NGS technology to detect AID mutations. NGS is being widely used for the study of clonal nucleotide changes, such as the detection of cancer-associated mutations (Ngo et al., 2011; Puente et al., 2011). However, AID initiated mutations occur at relatively low frequencies (10^{-5} to 10^{-3} mutations per base pair) and widespread in a given sequence, which makes depth coverage and fidelity critical for their detection. We have shown that the PCR-based NGS approach yields an average depth close to a hundred thousands of readings per base pair and that the fidelity of the technique allows the detection of mutations at frequencies below 10^{-4} per base pair. We believe this powerful approach will be extremely valuable for future SHM studies.

Here we have performed a thorough analysis of AID mutation spectrum by NGS. In contrast to previous analogous analyses, we are scoring only the mutations at the cytosine (or guanine, when in the opposite strand) residues contained in WRCY/RGYW hotspots. In addition, it is worth mentioning that in NIH-3T3, as in other cell lines, AID fails to promote mutations at A/T residues (second phase of SHM). Although the reasons for this G/C-only mutation pattern are not well understood, it most likely reflects the combined activity of AID and UNG and the absence of error-prone mismatch repair. We find that although WRCY/RGYW hotspots are

indeed preferred targets for AID, they display highly variable susceptibility to mutation -more than a 20-fold difference between TACC and AGCT, for instance. Interestingly, the four most mutated hotspots (AGCT, AGCT, AGCA and AGCC) show a preference for A and G nucleotides at -2 and -1 positions of the mutated C and for T at the +1 position of the mutated C, which results in AGCT being the most mutated combination at both strands. This result reveals a more stringent sequence preference for the introduction of mutations than previously recognized and reinforces AGCT as a primordial target for AID activity (Betz et al., 1993; Dorner et al., 1998), presumably in combination with error-prone processing by UNG, as discussed below. We find a striking similarity in the profile of hotspot mutation in B and non-B cells, which includes a highly variable mutation susceptibility of different WRCY/RGYW and a preferred targeting of AGCT, AGCT, AGCA and AGCC motifs. These data indicate a general mechanism of AID targeting at G/C residues, irrespective of the cell type and seemingly of the levels of AID expression. The latter implies that AID expression levels, while critical for the quantitative outcome of its activity (Sernandez et al., 2008; Takizawa et al., 2008), do not have a major qualitative impact on its sequence specificity. In addition, the severe block of transversions at G/C residues observed in NIH-3T3 cells upon UNG inhibition reveals that as in B cells (Di Noia et al., 2006; Rada et al., 2004), UNG activity in non-B cells is the only glycosylase activity responsible for the channeling of AID generated U:G mismatches into G/C transversions. Hence, this is a universal pathway for the generation of this type of AID-initiated mutations.

UNG normally removes uracil from DNA and initiates error-free base excision repair (BER), usually by short-patch repair that involves the activity of polymerase β (Hagen et al., 2006; Visnes et al., 2009). However, in the case of AID-generated U:G mismatches, UNG triggers an error-prone resolution of these lesions that leads to the generation of mutations (SHM) or

DNA double strand breaks (CSR). This non-canonical outcome of UNG activity remains one of the most puzzling issues in the field. Here we find that besides impairing transversions at G/C pairs, UNG inhibition in non-B cells gives rise to an increase load of mutations, which is in agreement with previous observations made on UNG-deficient primary B cells and in Ugi-expressing DT40 cells (Di Noia and Neuberger, 2002; Rada et al., 2002). This shows that not all of the U:G lesions generated by AID are subject to error-prone repair, but rather that a fraction of them are processed by a high fidelity repair pathway that is dependent on UNG activity. One possible explanation for this result would be that UNG levels might be limiting to ensure the error-free repair of all U:G lesions. We find that enforcing UNG activity in overexpression experiments does not significantly reduce the overall mutation frequency; however, it does shift the mutation pattern leading to an increase in the transversion and a decrease in the transition mutation frequencies, very much in the reverse fashion of the alterations observed after UNG inhibition. This result suggests that UNG overexpression enhances two alternative pathways that repair the initial U:G lesion: on one hand, faithful repair, that leads to the decrease of the fraction of lesions that would generate transitions through replication; and on the other, error-prone repair, which increases the frequency of transversion mutations, presumably coupled to translesion synthesis (TLS) activities. Therefore, UNG levels do not seem to specifically restrict the error-free processing of AID-induced deaminations.

These alternative outcomes of UNG activity prompted us to approach the sequence specificity of error-prone *versus* error-free repair of AID deaminations. Importantly, we find that not all of the WRCY/RGYW AID mutational hotspots are equally susceptible to both pathways and particularly that hotspots with a higher susceptibility to undergo mutation preferentially harbor a T at position +1 of the mutated C. This finding raises the question of how specific sequence motifs can influence the molecular pathways acting downstream of the glycosylation

reaction. It is tempting to speculate that particular sequence contexts could shift the dynamics of UNG-mediated glycosylation, allowing the specific recruitment of TLS polymerases to the abasic site and leading to transversion mutations. Regardless of the molecular mechanism responsible for this effect, this is to our knowledge the first evidence that sequence environment influences the outcome of UNG activity therefore contributing to the mutagenic resolution of AID induced U:G mismatches (Figure 7). This observation implies that UNG can modify the intrinsic targeting specificity of AID and could, at least partially, explain the discrepancies between *in vitro* and *in vivo* mutational hotspots (Pham et al., 2003; Rogozin and Kolchanov, 1992) and possibly the differential mutability of different genes in B cells lacking UNG (Liu et al., 2008).

In summary, our work unveils a new layer of regulation of the specificity of AID-induced mutations that is dependent of U:G mismatch resolution by UNG in a sequence dependent fashion. In addition, we have shown here striking similarities of mutational specificity in B and non-B cells, suggesting a universal mechanism for the generation of AID deaminations and their processing into mutations. These results open new perspectives in the understanding of AID function in antibody diversification and cancer development.

Materials and Methods

Cell cultures

293T and NIH-3T3 cell lines were cultured in 10% FCS DMEM medium supplemented with 1 μ M OHT when indicated. Mouse primary B cells were purified from the spleens of C57/BL6 mice by immunomagnetic depletion with anti-CD43 beads (Miltenyi Biotec) and were cultured in 50 μ M 2- β -Mercaptoethanol (Invitrogen), 10 mM Hepes (Invitrogen), 10 ng/ml IL4 (PeproTech), 25 μ g/ml LPS (Sigma-Aldrich) and 10% FCS RPMI medium.

Retroviral constructs

The AID-ER and AID^{E58Q}-ER retroviral vectors were described in (Ramiro et al., 2006). Ugi cDNA was kindly provided by Dr Samuel E. Bennett (Oregon State University) and was PCR-amplified using the oligonucleotides (forward) 5'-

GGATCCCGCCACCATGACAAATTTATCTGACATCATTGAAAAAG-3' and (reverse) 5'-

CTCGAGTTATAACATTTTAATTTTATTTTCTCCAT-3'. PCR product was cloned in *Bam*HI-*Xho*I sites

of pMX-PIE retroviral vector (Barreto et al., 2003), or in *Eco*RI site of pBABE-Hygro retroviral

vector (Cell Biolabs). The full length cDNA for the nuclear isoform of the mouse UNG (UNG2)

was obtained from the Integrated Molecular Analysis of Genomes and their Expression

(I.M.A.G.E.) consortium (I.M.A.G.E. clone 4009947). The coding region was PCR-amplified using

the oligonucleotides (forward) 5'-AAGGATCCACCATGATCGGCCAGAAGACCCT-3' and (reverse)

5'-AAGGATCCGTCACAGCTCCTTCCAGTTGA-3' and cloned using pGEM-T easy vector kit

(Promega). UNG2 was excised from this vector using flanking *Eco*RI sites and subcloned into

pBABE-Hygro retroviral vector. mOrange point mutations were introduced by standard site-

directed mutagenesis to create the STOP codon, using the oligonucleotides (forward) 5'-

AAGGATCCCCACCATGGTGAGCAAGGGCGAGGAGAATAACA-3', (reverse) 5'-

TGTCGGCGGGGTGCTTCAGCTAGGCCTTGGAGCC-3', (forward) 5'-

GGCTCCAAGGCCTAGCTGAAGCACCCCGCCGACA-3' and (reverse) 5'-

AAGGATCCTTACTTGTACAGCTCGTCCATGC-3' and cloned using pGEM-T easy vector kit.

mOrange^{STOP} was excised from this vector using flanking *Bam*HI sites and subcloned into pMX-PIE retroviral vector (Barreto et al., 2003).

Retroviral infection and flow cytometry analysis

Retroviral supernatants were produced by transient calcium phosphate cotransfection with pCL-ECO (Imgenex) and a combination of AID-ER, AID^{E58Q}-ER, pMX-PIE-Orange^{STOP}, pBABE-Ugi, pBABE-UNG or pMX-PIE-Ugi retroviral vectors. NIH-3T3 cells were transduced with retroviral supernatants for 16 h in the presence of 8 µg/ml polybrene (Sigma-Aldrich) and selected with 0.4 µg/ml puromycin or 100 µg/ml hygromycin. Transduced NIH-3T3 were cultured in the presence of 1 µM OHT and GFP⁺ mOrange⁺ cells were monitored by flow cytometry (FACSCanto; BD Biosciences) at the indicated time points. Mouse primary B cells were transduced with retroviral supernatants for 20 h in the presence of 8 µg/ml polybrene, 25 µg/ml LPS and 10 ng/ml IL4. 24, 48 and 72 h after transduction, primary B cells were labeled with anti-IgG1 biotin antibody and APC-streptavidin (BD biosciences) to determine IgG1 expression in GFP⁺ cells by flow cytometry.

mOrange^{STOP} amplification and mutation analysis

For analysis of mutations at mOrange^{STOP} gene, co-transduced NIH-3T3 cells were cultured for 11 days in the presence of OHT. DNA was extracted and 1x10⁴ cells were amplified using the oligonucleotides (forward) 5'- AATAACATGGCCATCATCAAGGA-3' and (reverse) 5'- ACGTAGCGGCCCTCGAGTCTCTC-3' with 2.5 U of Pfu Ultra (Stratagene) in a 50 µl reaction under the following conditions: 94°C for 5 min followed by 25 cycles at 94°C for 10 s, 60°C for 30 s and 72°C for 1 min. For Sanger sequencing, PCR products were cloned using pGEM-T easy

vector kit and plasmids from individual colonies were sequenced using SP6 universal primer. Sequence analysis was performed using SeqMan software (Lasergene). For NGS sequencing, 5 independent PCR reactions were pooled and equimolar amounts of each sample were mixed and fragmented to a range of 100-300pb (Covaris S2). DNA was processed through successive enzymatic treatments of end-repair, dA-tailing, and ligation to adapters following manufacturer's instructions (Illumina). Adapter-ligated libraries were further amplified by PCR with Phusion High-Fidelity DNA Polymerase (Finnzymes) using Illumina PE primers for 10 cycles. The resulting purified DNA libraries were multiplexed and applied to an Illumina flow cell for cluster generation, and sequenced on the Genome Analyzer IIx with SBS TruSeq v5 reagents following manufacturer's protocols. High quality reads (Phred Quality Score>20) were aligned using Novoalign (Novocraft technologies) and were processed with SAMtools (Li et al., 2009) and custom scripting.

S μ amplification and mutation analysis

For analysis of mutations at S μ region, mouse primary B cells from the spleens of wild type and AID knock-out mice (n=2) were purified and cultured for 3 days in the presence of LPS and IL4 as described above. DNA was extracted and S μ was PCR-amplified using the oligonucleotides (forward) 5'- AATGGATACCTCAGTGGTTTTTAATGGTGGGTTTA-3' and (reverse) 5'- GCGGCCCGGCTCATTCCAGTTCATTACAG-3'. For each sample, 1.3x10⁴ cells were amplified using Pfu Ultra for 25 cycles (94°C for 30 s, 60°C for 30 s and 72°C for 50 s). 8 independent PCR amplification reactions were pooled, processed, sequenced by NGS and analyzed as described for mOrange^{STOP} gene.

Statistical analysis

In vitro CSR and mOrange⁺ cells assays, and mutation frequencies were analyzed with Student's t tests. Mutation percentages were analyzed with Fisher's exact test.

Acknowledgements

We would like to thank Vasco Barreto, Luis Blanco, Oskar Fernández-Capetillo and Juan Méndez for critical reading of the manuscript, Samuel E. Bennett for kindly providing the PSB2 cDNA and O Domínguez for technical advice. PPD is a fellow of the research training program (FPI) funded by the Ministerio de Ciencia e Innovación, LB and ARR are supported by the Spanish National Cancer Research Center (CNIO), VGY is a Ramón y Cajal Investigator (Ministerio de Ciencia e Innovación) and PD is funded by the European Research Council Starting Grant program (BCLYM-207844). This work was funded by grants from Ministerio de Ciencia e Innovación (SAF2010-21394) and European Research Council Starting Grant program (BCLYM-207844).

References

- Adams, J.M., A.W. Harris, C.A. Pinkert, L.M. Corcoran, W.S. Alexander, S. Cory, R.D. Palmiter, and R.L. Brinster. 1985. The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature*. 318:533-8.
- Bachl, J., and C. Olsson. 1999. Hypermutation targets a green fluorescent protein-encoding transgene in the presence of immunoglobulin enhancers. *Eur J Immunol*. 29:1383-9.
- Barreto, V., B. Reina-San-Martin, A.R. Ramiro, K.M. McBride, and M.C. Nussenzweig. 2003. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell*. 12:501-8.
- Betz, A.G., M.S. Neuberger, and C. Milstein. 1993. Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin V genes. *Immunol Today*. 14:405-11.
- Bransteitter, R., P. Pham, P. Calabrese, and M.F. Goodman. 2004. Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. *J Biol Chem*. 279:51612-21.
- Di Noia, J., and M.S. Neuberger. 2002. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature*. 419:43-8.
- Di Noia, J.M., and M.S. Neuberger. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem*. 76:1-22.
- Di Noia, J.M., C. Rada, and M.S. Neuberger. 2006. SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J*. 25:585-95.
- Doi, T., K. Kinoshita, M. Ikegawa, M. Muramatsu, and T. Honjo. 2003. De novo protein synthesis is required for the activation-induced cytidine deaminase function in class-switch recombination. *Proc Natl Acad Sci U S A*. 100:2634-8.
- Dorner, T., S.J. Foster, N.L. Farner, and P.E. Lipsky. 1998. Somatic hypermutation of human immunoglobulin heavy chain genes: targeting of RGYW motifs on both DNA strands. *Eur J Immunol*. 28:3384-96.
- Dorsett, Y., D.F. Robbiani, M. Jankovic, B. Reina-San-Martin, T.R. Eisenreich, and M.C. Nussenzweig. 2007. A role for AID in chromosome translocations between c-myc and the IgH variable region. *J Exp Med*. 204:2225-32.
- Endo, Y., H. Marusawa, K. Kinoshita, T. Morisawa, T. Sakurai, I.M. Okazaki, K. Watashi, K. Shimotohno, T. Honjo, and T. Chiba. 2007. Expression of activation-induced cytidine deaminase in human hepatocytes via NF-kappaB signaling. *Oncogene*. 26:5587-95.
- Hagen, L., J. Pena-Diaz, B. Kavli, M. Otterlei, G. Slupphaug, and H.E. Krokan. 2006. Genomic uracil and human disease. *Exp Cell Res*. 312:2666-72.
- Imai, K., G. Slupphaug, W.I. Lee, P. Revy, S. Nonoyama, N. Catalan, L. Yel, M. Forveille, B. Kavli, H.E. Krokan, H.D. Ochs, A. Fischer, and A. Durandy. 2003. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol*. 4:1023-8.
- Kou, T., H. Marusawa, K. Kinoshita, Y. Endo, I.M. Okazaki, Y. Ueda, Y. Kodama, H. Haga, I. Ikai, and T. Chiba. 2007. Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. *Int J Cancer*. 120:469-76.
- Kovalchuk, A.L., W. duBois, E. Mushinski, N.E. McNeil, C. Hirt, C.F. Qi, Z. Li, S. Janz, T. Honjo, M. Muramatsu, T. Ried, T. Behrens, and M. Potter. 2007. AID-deficient Bcl-xL transgenic mice develop delayed atypical plasma cell tumors with unusual Ig/Myc chromosomal rearrangements. *J Exp Med*. 204:2989-3001.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078-9.

- Liu, M., J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, and D.G. Schatz. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature*. 451:841-5.
- Martin, A., and M.D. Scharff. 2002. Somatic hypermutation of the AID transgene in B and non-B cells. *Proc Natl Acad Sci U S A*. 99:12304-8.
- Matsumoto, Y., H. Marusawa, K. Kinoshita, Y. Endo, T. Kou, T. Morisawa, T. Azuma, I.M. Okazaki, T. Honjo, and T. Chiba. 2007. Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med*. 13:470-6.
- Maul, R.W., H. Saribasak, S.A. Martomo, R.L. McClure, W. Yang, A. Vaisman, H.S. Gramlich, D.G. Schatz, R. Woodgate, D.M. Wilson, 3rd, and P.J. Gearhart. 2011. Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat Immunol*. 12:70-6.
- McBride, K.M., V. Barreto, A.R. Ramiro, P. Stavropoulos, and M.C. Nussenzweig. 2004. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J Exp Med*. 199:1235-44.
- Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*. 102:553-63.
- Ngo, V.N., R.M. Young, R. Schmitz, S. Jhavar, W. Xiao, K.H. Lim, H. Kohlhammer, W. Xu, Y. Yang, H. Zhao, A.L. Shaffer, P. Romesser, G. Wright, J. Powell, A. Rosenwald, H.K. Muller-Hermelink, G. Ott, R.D. Gascoyne, J.M. Connors, L.M. Rimsza, E. Campo, E.S. Jaffe, J. Delabie, E.B. Smeland, R.I. Fisher, R.M. Braziel, R.R. Tubbs, J.R. Cook, D.D. Weisenburger, W.C. Chan, and L.M. Staudt. 2011. Oncogenically active MYD88 mutations in human lymphoma. *Nature*. 470:115-9.
- Pasqualucci, L., G. Bhagat, M. Jankovic, M. Compagno, P. Smith, M. Muramatsu, T. Honjo, H.C. Morse, 3rd, M.C. Nussenzweig, and R. Dalla-Favera. 2008. AID is required for germinal center-derived lymphomagenesis. *Nat Genet*. 40:108-12.
- Pasqualucci, L., P. Neumeister, T. Goossens, G. Nanjangud, R.S. Chaganti, R. Kuppers, and R. Dalla-Favera. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*. 412:341-6.
- Peled, J.U., F.L. Kuang, M.D. Iglesias-Ussel, S. Roa, S.L. Kalis, M.F. Goodman, and M.D. Scharff. 2008. The biochemistry of somatic hypermutation. *Annu Rev Immunol*. 26:481-511.
- Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature*. 418:99-103.
- Pham, P., R. Bransteitter, J. Petruska, and M.F. Goodman. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*. 424:103-7.
- Puente, X.S., M. Pinyol, V. Quesada, L. Conde, G.R. Ordonez, N. Villamor, G. Escaramis, P. Jares, S. Bea, M. Gonzalez-Diaz, L. Bassaganyas, T. Baumann, M. Juan, M. Lopez-Guerra, D. Colomer, J.M. Tubio, C. Lopez, A. Navarro, C. Tornador, M. Aymerich, M. Rozman, J.M. Hernandez, D.A. Puente, J.M. Freije, G. Velasco, A. Gutierrez-Fernandez, D. Costa, A. Carrio, S. Guijarro, A. Enjuanes, L. Hernandez, J. Yague, P. Nicolas, C.M. Romeo-Casabona, H. Himmelbauer, E. Castillo, J.C. Dohm, S. de Sanjose, M.A. Piris, E. de Alava, J. San Miguel, R. Royo, J.L. Gelpi, D. Torrents, M. Orozco, D.G. Pisano, A. Valencia, R. Guigo, M. Bayes, S. Heath, M. Gut, P. Klatt, J. Marshall, K. Raine, L.A. Stebbings, P.A. Futreal, M.R. Stratton, P.J. Campbell, I. Gut, A. Lopez-Guillermo, X. Estivill, E. Montserrat, C. Lopez-Otin, and E. Campo. 2011. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 475:101-5.

- Rada, C., J.M. Di Noia, and M.S. Neuberger. 2004. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell*. 16:163-71.
- Rada, C., M.R. Ehrenstein, M.S. Neuberger, and C. Milstein. 1998. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity*. 9:135-41.
- Rada, C., G.T. Williams, H. Nilsen, D.E. Barnes, T. Lindahl, and M.S. Neuberger. 2002. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol*. 12:1748-55.
- Ramiro, A.R., M. Jankovic, E. Callen, S. Difilippantonio, H.T. Chen, K.M. McBride, T.R. Eisenreich, J. Chen, R.A. Dickins, S.W. Lowe, A. Nussenzweig, and M.C. Nussenzweig. 2006. Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. *Nature*. 440:105-9.
- Ramiro, A.R., M. Jankovic, T. Eisenreich, S. Difilippantonio, S. Chen-Kiang, M. Muramatsu, T. Honjo, A. Nussenzweig, and M.C. Nussenzweig. 2004. AID is required for c-myc/IgH chromosome translocations in vivo. *Cell*. 118:431-8.
- Revy, P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labeu, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A.G. Ugazio, N. Brousse, M. Muramatsu, L.D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer, and A. Durandy. 2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell*. 102:565-75.
- Robbiani, D.F., A. Bothmer, E. Callen, B. Reina-San-Martin, Y. Dorsett, S. Difilippantonio, D.J. Bolland, H.T. Chen, A.E. Corcoran, A. Nussenzweig, and M.C. Nussenzweig. 2008. AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell*. 135:1028-38.
- Rogozin, I.B., and N.A. Kolchanov. 1992. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim Biophys Acta*. 1171:11-8.
- Schrader, C.E., W. Edelmann, R. Kucherlapati, and J. Stavnezer. 1999. Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. *J Exp Med*. 190:323-30.
- Sernandez, I.V., V.G. de Yebenes, Y. Dorsett, and A.R. Ramiro. 2008. Haploinsufficiency of activation-induced deaminase for antibody diversification and chromosome translocations both in vitro and in vivo. *PLoS One*. 3:e3927.
- Shaner, N.C., R.E. Campbell, P.A. Steinbach, B.N. Giepmans, A.E. Palmer, and R.Y. Tsien. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol*. 22:1567-72.
- Shen, H.M., A. Peters, B. Baron, X. Zhu, and U. Storb. 1998. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science*. 280:1750-2.
- Stavnezer, J., J.E. Guikema, and C.E. Schrader. 2008. Mechanism and regulation of class switch recombination. *Annu Rev Immunol*. 26:261-92.
- Takai, A., H. Marusawa, Y. Minaki, T. Watanabe, H. Nakase, K. Kinoshita, G. Tsujimoto, and T. Chiba. 2011. Targeting activation-induced cytidine deaminase prevents colon cancer development despite persistent colonic inflammation. *Oncogene*.
- Takizawa, M., H. Tolarova, Z. Li, W. Dubois, S. Lim, E. Callen, S. Franco, M. Mosaico, L. Feigenbaum, F.W. Alt, A. Nussenzweig, M. Potter, and R. Casellas. 2008. AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *J Exp Med*. 205:1949-57.

- Visnes, T., B. Doseth, H.S. Pettersen, L. Hagen, M.M. Sousa, M. Akbari, M. Otterlei, B. Kavli, G. Slupphaug, and H.E. Krokan. 2009. Uracil in DNA and its processing by different DNA glycosylases. *Philos Trans R Soc Lond B Biol Sci.* 364:563-8.
- Wang, Z., and D.W. Mosbaugh. 1988. Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in *Escherichia coli*. *J Bacteriol.* 170:1082-91.
- Wiesendanger, M., B. Kneitz, W. Edelmann, and M.D. Scharff. 2000. Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J Exp Med.* 191:579-84.
- Xue, K., C. Rada, and M.S. Neuberger. 2006. The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in *msh2*^{-/-} *ung*^{-/-} mice. *J Exp Med.* 203:2085-94.
- Yoshikawa, K., I.M. Okazaki, T. Eto, K. Kinoshita, M. Muramatsu, H. Nagaoka, and T. Honjo. 2002. AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science.* 296:2033-6.
- Yu, K., F.T. Huang, and M.R. Lieber. 2004. DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J Biol Chem.* 279:6496-500.
- Zarrin, A.A., F.W. Alt, J. Chaudhuri, N. Stokes, D. Kaushal, L. Du Pasquier, and M. Tian. 2004. An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nat Immunol.* 5:1275-81.

Figure Legends

Figure 1. Fluorescence revertance assay to monitor AID activity

A) Representation of the pMX-PIE-mOrange^{STOP} and AID-ER-ΔhuCD4 or AID^{E58Q}-ER-ΔhuCD4 retroviral vectors used to transduce NIH-3T3 cells. B) Detection of AID activity. NIH-3T3 cells were co-transduced with retroviral vectors depicted in panel A and cultured with (bottom panels) or without (upper panels) 1 μM OHT. mOrange⁺ cells were monitored by flow cytometry. Representative analyses at day 9 are shown. C) Time-course analysis of mOrange⁺ cell appearance in NIH-3T3 cells co-transduced as described in panel B.

Figure 2. UNG inhibition by Ugi blocks CSR and the generation of mOrange⁺ cells

A) Ugi expression blocks CSR. Primary B cells were isolated from spleens of wild type mice, cultured in the presence of LPS and IL4 and transduced with Ugi or empty vector as a control. CSR efficiency was measured by flow cytometry analysis of IgG1 expression 3 days after retroviral transduction. Representative analyses are shown. B) Time-course analysis of IgG1 CSR efficiency of control and Ugi-transduced B cells in LPS + IL4 cultures, relative to control IgG1 expression at day 2. Means from 3-7 experiments are shown. Error bars show the standard deviation values at each time point (p<0.01 at all days). C) Ugi expression blocks the generation of mOrange⁺ cells. NIH-3T3 cells were co-transduced with mOrange^{STOP} and AID-ER expressing vectors, and Ugi or empty vector as control. Transduced cells were cultured in the presence of OHT and percentage of mOrange⁺ cells was monitored by flow cytometry. Representative analyses after 9 days of OHT treatment are shown. D) Time-course analysis of mOrange⁺ cell appearance in co-transduced NIH-3T3, relative to control mOrange⁺ cells after 2 days of OHT treatment. Means from 3-7 experiments are represented (p<0.01 at all days).

Figure 3. UNG inhibition by Ugi increases global mutation frequency in mOrange^{STOP}

NIH-3T3 cells were co-transduced with mOrange^{STOP} and AID-ER expressing vectors, and Ugi or empty vector as control. Transduced cells were cultured in the presence of OHT during 11 days, and mOrange^{STOP} gene was cloned and sequenced by standard Sanger method. A) Pattern of nucleotide substitutions in mOrange^{STOP} gene in control (upper panel) and Ugi-expressing (bottom panel) NIH-3T3 cells. Mutation numbers for each nucleotide substitution are shown. B) Relative contribution of G/C transversions, G/C transitions and A/T mutations. Values are expressed as the percentage of total mutations (G/C transversions, $p=3.5 \times 10^{-7}$; G/C transitions, $p=2.8 \times 10^{-8}$; A/T mutations, $p=0.06$). C) Absolute frequency of G/C transversions, G/C transitions and A/T mutations (G/C transversions, $p=0.0084$; G/C transitions, $p=0.0026$; A/T mutations, $p=0.29$).

Figure 4. Analysis of AID-induced mutations by next generation sequencing

NIH-3T3 cells were co-transduced with mOrange^{STOP} and AID-ER or AID^{E58Q}-ER expressing vectors. Transduced cells were cultured in the presence of OHT during 11 days. A) mOrange^{STOP} gene was cloned and sequenced by standard Sanger method. Mutation frequency in AID-ER expressing cells in the complete mOrange^{STOP} sequence (Total) and in cytidines or guanines of AID hotspots (WRC/GYW and WRCY/RGYW) is represented. B) mOrange^{STOP} gene was PCR-amplified from AID-ER and AID^{E58Q}-ER expressing cells and sequenced by NGS. Mutation frequency was calculated as mean frequency per base-pair in AID-ER (black bars) or AID^{E58Q}-ER (white bars) expressing cells and is represented as described in panel A. C) Mutation frequency in each of the WRCY and RGYW hotspots in mOrange^{STOP} sequence, calculated as described in panel B after subtracting AID^{E58Q}-ER values per position. Bars represent the mean of 3 independent experiments. Red dotted line represents the mean mutation frequency at G/C pairs in mOrange^{STOP} sequence. Grey shadow indicates that the particular hotspot is not present in mOrange^{STOP} sequence. D) NGS analysis in μ switch (S_{μ}) region. Primary spleen B

cells from AID^{+/+} and AID^{-/-} mice were isolated and cultured in the presence of LPS and IL4 for 72 h. $\Sigma\mu$ region from two different animals per genotype was PCR-amplified and sequenced by NGS. Mutation frequency was calculated as the mean of both animals per base-pair after subtracting AID^{-/-} frequencies, and is represented as described in panel C.

Figure 5. UNG overexpression increases mOrange⁺ revertants and modifies the resolution of AID-induced lesions

NIH-3T3 cells were co-transduced with mOrange^{STOP} and AID-ER expressing vectors, and Ugi, UNG or empty vector as control. A) UNG overexpression increases the frequency of mOrange⁺ cells. Transduced cells were cultured in the presence of OHT during 11 days and percentage of mOrange⁺ cells was monitored by flow cytometry. Representative analyses at day 9 are shown. B) Time-course analysis of mOrange⁺ cells appearance in co-transduced NIH-3T3, relative to control mOrange⁺ cells after 2 days of OHT treatment. Means from 3-7 experiments are represented ($p < 0.05$ at all days). C) mOrange^{STOP} gene was PCR- amplified from co-transduced NIH-3T3 and sequenced by NGS. Relative contribution to total mutation frequency of G/C transversions, G/C transitions and A/T mutations is represented. Bars show the mean values of three different experiments. D) Absolute frequency of G/C transversions, G/C transitions and A/T mutations. Bars show the mean values of three different experiments, calculated as mean frequency per base-pair.

Figure 6. UNG displays sequence preference to promote error-free or high fidelity repair at AID mutational hotspots

Mutation analysis in mOrange^{STOP} sequence of NIH-3T3 cells co-transduced with mOrange^{STOP} and AID-ER expressing vectors, and Ugi (white bars), UNG (grey bars) or empty vector as control (black bars). A) Relative transversion frequency at G²³¹ is represented. Bars show the

mean values of three different experiments (UNG vs control p (paired *t* test)=0.032). B) G/C transition (upper panel) and G/C transversion (bottom panel) frequency in each of the WRCY and RGYW motifs in the mOrange^{STOP} sequence, calculated as mean frequency per base-pair (n=3). C) Ratio of G/C transition frequency between Ugi expressing and control cells (upper panel) and of G/C transversion frequency between UNG expressing and control cells (bottom panel) (n=3).

Figure 7. UNG shapes the specificity of AID-induced mutations

Diagram summarizes the main findings of this work (see text for details). Alternative processing pathways leading to the generation of DSBs or A/T mutations are purposely excluded from the model for the sake of clarity.

Table I. mOrangeSTOP sequencing by Sanger or next generation sequencing

	Sanger	NGS
Number of positions^a	222	222
Number of reads (mean)^b	101	87009
Mutation number^c	31	22370
Frequency (x10⁻⁴)^d	13.8	11.1

^aNumber of total cytosines in mOrange sequence. ^bMean number of reads per cytosine. ^cTotal mutation number in cytosines. ^dMutation frequency calculated as $c/(b \times a)$

Figure 1

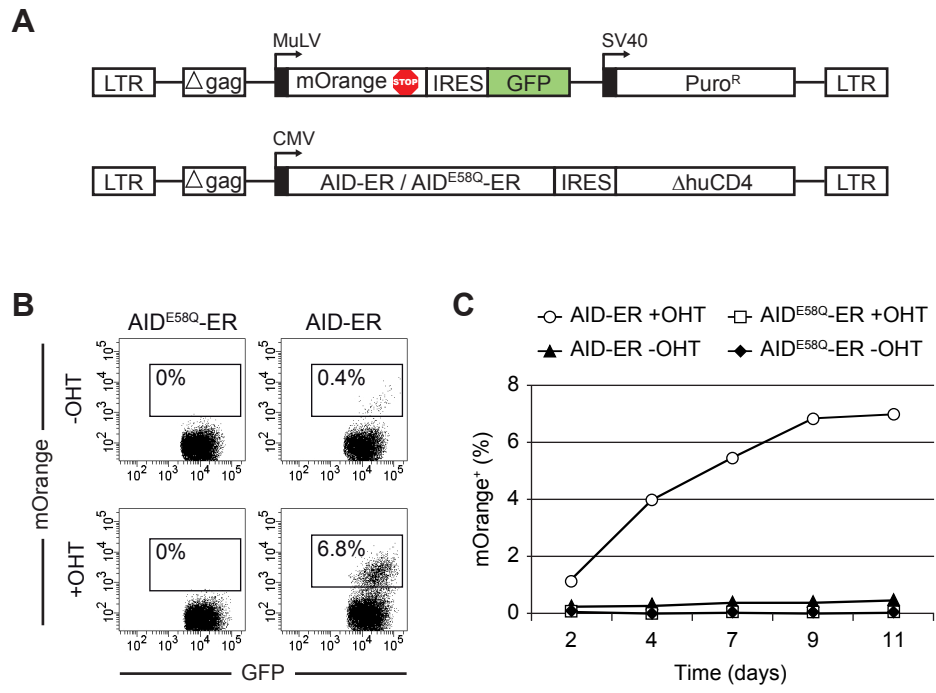


Figure 2

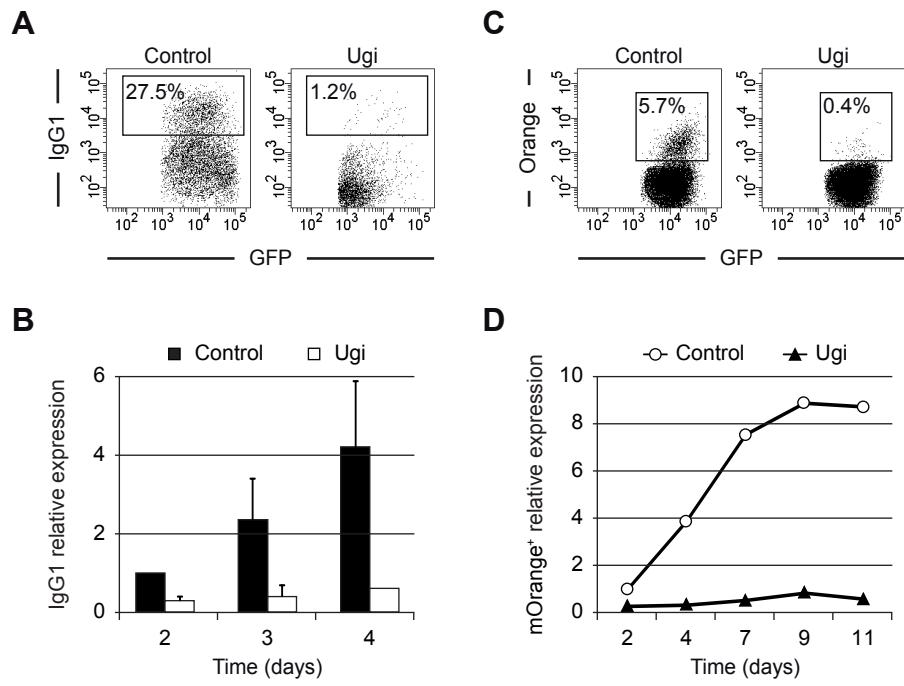


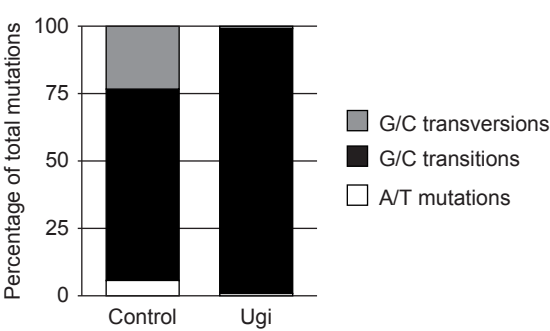
Figure 3

A

		From →			
		A	C	G	T
To ↓	A		2	24	1
	C	1		6	-
	G	-	4		1
	T	1	25	4	
Total		2	31	34	2

		From →			
		A	C	G	T
To ↓	A		1	81	-
	C	1		-	-
	G	-	-		-
	T	-	37	-	
Total		1	38	81	0

B



C

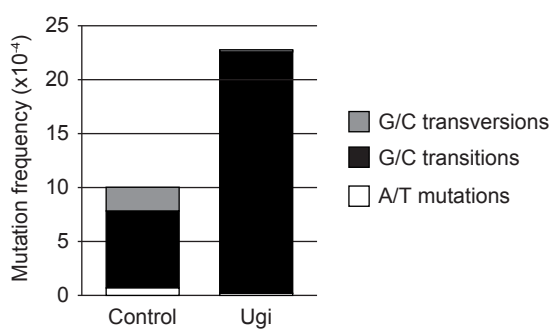


Figure 4

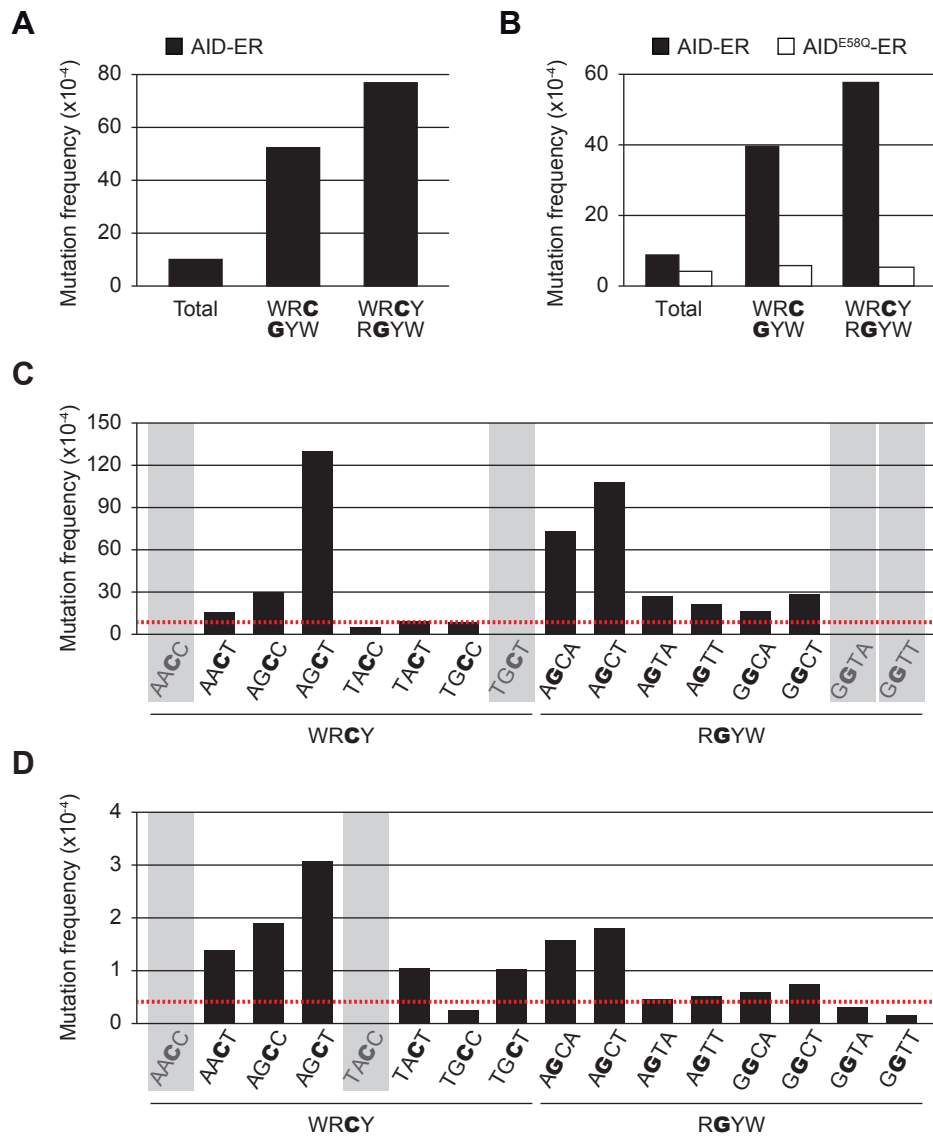
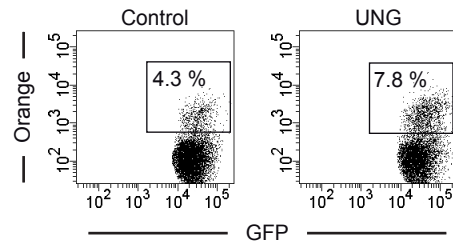
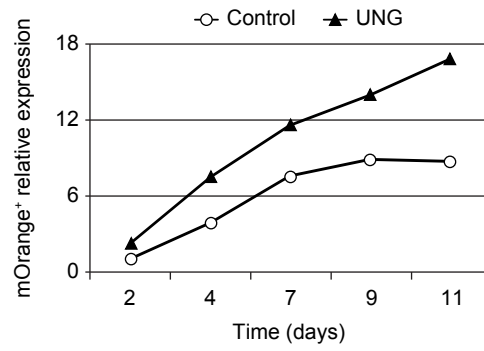


Figure 5

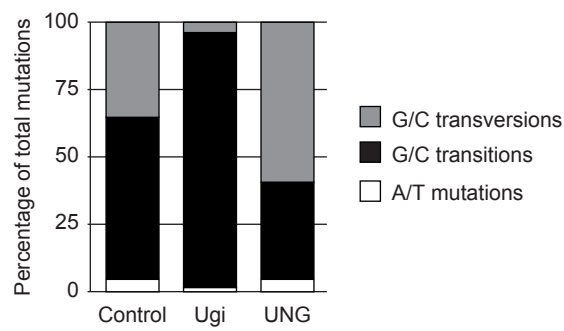
A



B



C



D

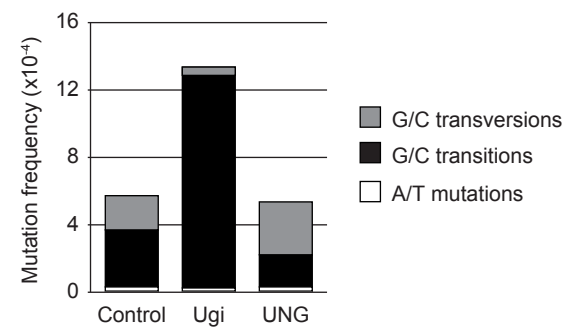


Figure 6

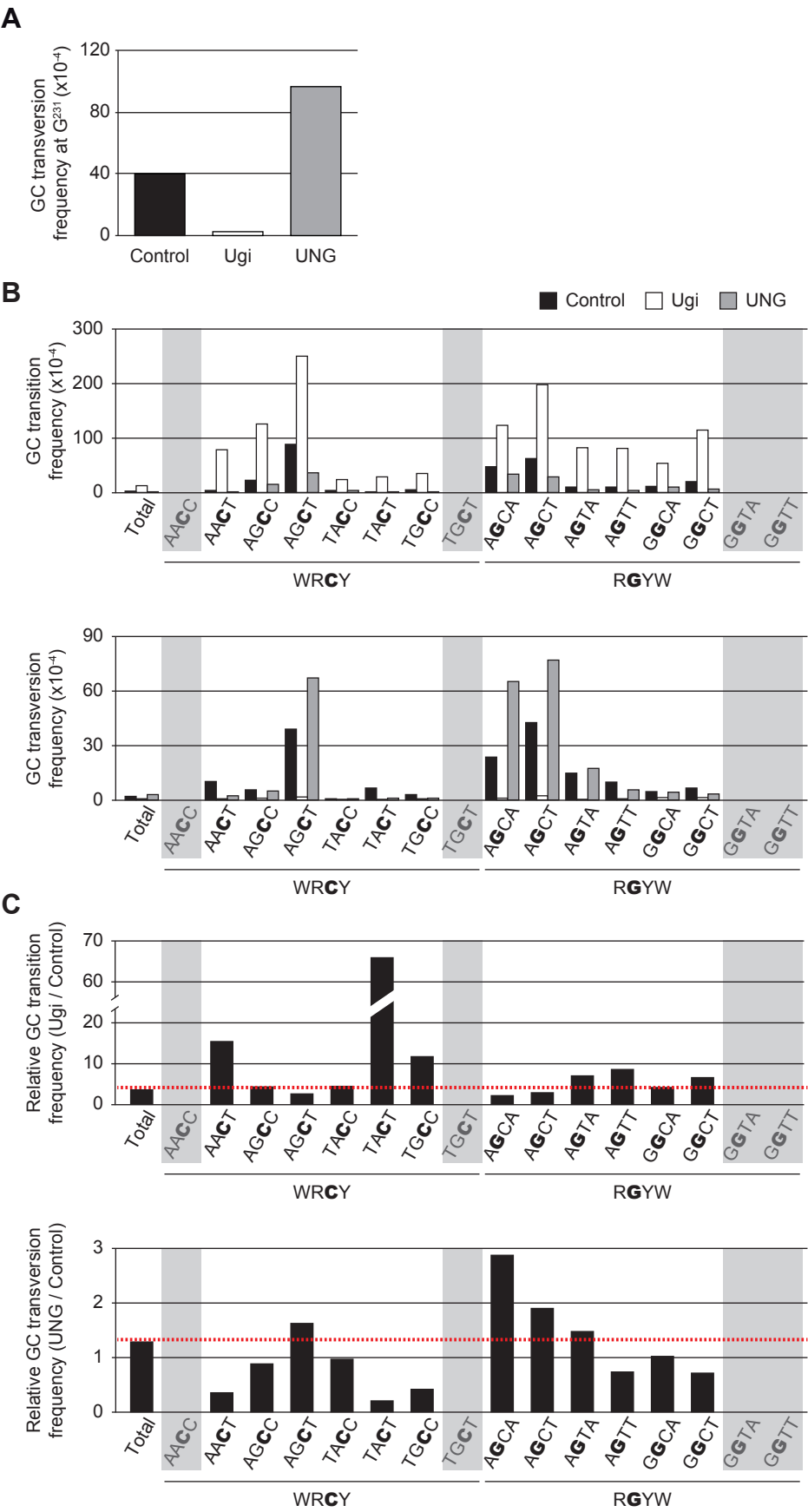
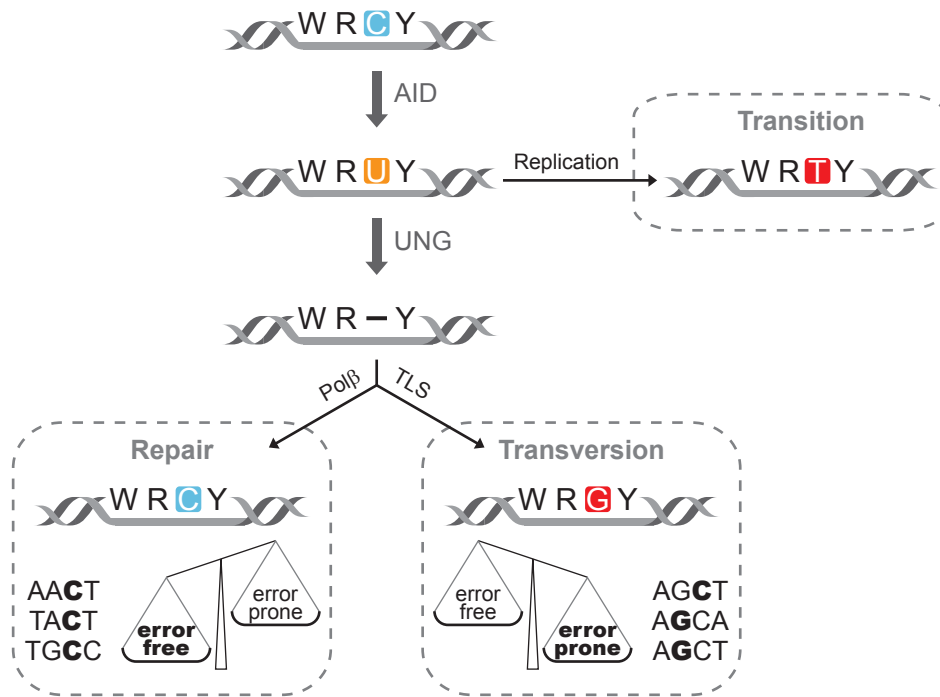


Figure 7



Oncogenic events triggered by AID, the adverse effect of antibody diversification

Pablo Pérez-Durán, Virginia G.de Yebenes and Almudena R.Ramiro*

DNA Hypermutation and Cancer group, Spanish National Research Cancer Center, Melchor Fernandez Almagro 3, 28029 Madrid, Spain

*To whom correspondence should be addressed. Tel: +34 91 224 6900;

Fax: +34 91 732 8033;

Email: aramiro@cniio.es

The generation of an efficient immune response depends on highly refined mechanisms of antibody diversification. Two of these mechanisms, somatic hypermutation (SHM) and class switch recombination (CSR), are initiated by activation-induced cytidine deaminase (AID) upon antigen stimulation of mature B cells. AID deaminates cytosines on the DNA of Ig genes thereby generating a lesion that can be processed into a mutation (SHM) or a DNA double-strand break followed by a recombination reaction (CSR). A number of mechanisms are probably responsible for regulating AID function, such as transcriptional regulation, subcellular localization, post-transcriptional modifications and target specificity, but the issue remains of how unwanted DNA damage is fully prevented. Most lymphocyte neoplasias are originated from mature B cells and harbour hallmark chromosome translocations of lymphomagenic potential, such as the c-myc/IgH translocations found in Burkitt lymphomas. It has been recently shown that such translocations are initiated by AID and that ataxia-telangiectasia mutated, p53 and ARF provide surveillance mechanisms to prevent these aberrations. In addition, evidence is accumulating that AID expression can be induced in B cells independently of the germinal centre environment, such as in response to some viral infections, and occasionally in non-B cells, at least in certain inflammation-associated neoplastic situations. The most recent findings on AID expression and function and their relevance to the generation of oncogenic lesions will be discussed.

Introduction

Adaptive immunity is an exquisite evolutionary development of vertebrates which provides a highly specialized defence barrier against pathogens. The hallmark of adaptive immunity is its extreme specificity for foreign antigens, enabled by antigen receptors of virtually unlimited diversity. Highly diverse antigen receptors are generated during the differentiation of T and B lymphocytes through somatic recombination of their T-cell receptor and immunoglobulin genes, respectively. Germ line antigen receptor genes comprise a collection of variable (V), diverse (D) and joining (J) segments. A combination of V, D and J segment is randomly assembled in each lymphocyte by the site-specific V(D)J recombination reaction giving rise to the antigen recognition site of the receptor. Therefore, a plethora of different antigen receptors is generated by combinatorial diversity. In contrast to T lymphocytes, mature B cells have as yet another chance to further diversify their receptors after antigen encounter. This refinement takes place in germinal centres through two independent mechanisms, somatic hypermutation (SHM) and class switch recombination (CSR).

Abbreviations: ATM, ataxia-telangiectasia mutated; Ab-MLV, Abelson murine leukaemia virus; AID, activation-induced deaminase; AIDtg, AID transgenic; CSR, class switch recombination; DSB, double-strand break; EBNA, Epstein–Barr virus nuclear antigen; EBV, Epstein–Barr virus; HCV, hepatitis C virus; HCC, hepatocellular carcinoma; IgH, immunoglobulin heavy chain; LMP-1, latent membrane protein 1; NF-κB, nuclear factor-kappa B; SHM, somatic hypermutation; ssDNA, single-stranded DNA; UNG, uracil-N-glycosylase; V(D)J, variable, diverse and joining segments.

The Ig molecule is composed of two identical pairs of light chains (IgL) and heavy chains (IgH). The N-terminal portions of IgL and IgH comprise the antigen-binding site, coded by rearranged V(D)J variable segments. The C-terminal region of IgH is called constant region and determines the immunological pathway leading to antigen removal. The IgH locus is organized in a number of constant regions, each of which is preceded by a highly repetitive, non-coding sequence called switch region (Figure 1). Before antigen encounter, naive B cells express Igs bearing the most upstream μ (IgM) or δ (IgD) constant regions. Antigen recognition by naive B cells can trigger the germinal centre reaction leading to the secondary diversification of the immunoglobulin genes through CSR and SHM reactions. This secondary reshaping is crucial to improve the efficiency and affinity of the antibody response. SHM involves the introduction of nucleotide changes, usually point mutations, on the productively rearranged V(D)J segments that code for the variable region of the antibody molecule. B cells whose Ig genes acquire mutations that improve the receptor affinity for antigen will proliferate and differentiate into either antibody-producing plasma cells or memory B cells; cells where SHM results in low affinity or deleterious mutations in their Ig genes will be rendered anergic. CSR is a region-specific recombination reaction that takes place between two switch regions of the IgH locus thereby replacing the C μ constant region by a downstream constant region that will endow the Ig with different effector functions (Figure 1). Both SHM and CSR are initiated by the same B-cell-specific enzymatic activity, activation-induced cytidine deaminase (AID), through the deamination of cytosines on the variable or switch regions of the Ig locus, respectively (Figure 1). This initial lesion on DNA is subsequently processed to allow the fixation of a mutation, in the case of SHM, or a recombination reaction, in the case of CSR.

About 95% of human lymphomas arise from B cells, rather than T cells. Among those, most have their origin in mature—germinal centre or post-germinal centre—B cells. The hallmark of these malignancies is the presence of lymphomagenic chromosomal translocations that involve the Ig locus and a proto-oncogene (reviewed in ref. 1). Evidence has accumulated to prove the involvement of the germinal centre reactions and specifically of AID activity in the generation of lymphoma-associated translocations. Furthermore, newest results show that AID expression might not be absolutely restricted to germinal centre B cells and that this expression could in turn have implications in the development of neoplasias other than B-cell lymphomas.

AID in antibody diversification

AID function. AID was identified by the Honjo lab in 1999 (2) in a search for cDNAs that were specifically expressed in B cells upon CSR activation. Soon after that generation of AID-deficient mice proved that AID is required not only for CSR but also for SHM (3), and that mutations in the AID gene in humans are associated with an immune deficiency called Hyper-IgM syndrome type 2 (4). Moreover, AID expression in fibroblasts transfected with appropriate substrates can reconstitute both SHM and CSR (5,6), thus proving that AID is the only B-cell-specific factor required to trigger the germinal centre reactions.

The actual mechanism of AID activity in promoting these reactions has been the subject of intense debate (reviewed in ref. 7) regarding the nature of its substrate—RNA or DNA. Today, there is compelling genetic and biochemical evidence showing that AID initiates SHM and CSR by deaminating cytosines on the DNA of Ig variable or switch regions, respectively. Cytosine deamination results in the generation of a uracil and therefore turns a DNA C:G pair into a U:G mismatch. This AID-generated U:G mismatch can undergo different fates: (i) it can be replicated over, giving rise to transition mutations at G:C pairs; (ii) it can be recognized by mismatch repair factors

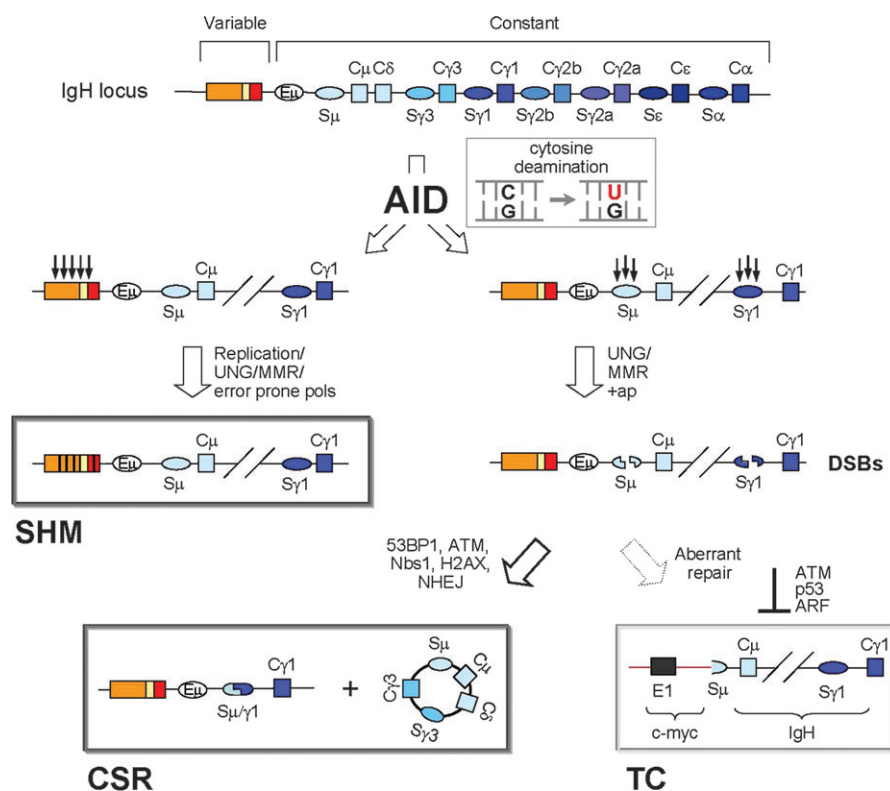


Fig. 1. Molecular events triggered by AID. Mature B cells harbour IgH genes (top diagram) that comprise a 5' portion that has been productively rearranged by V(D)J recombination (orange, yellow and red boxes) coding for the antigen recognition variable region of the antibody molecule, and a number of constant regions (blue boxes), preceded by highly repetitive switch regions (blue ovals). Antibody diversification by SHM and CSR is triggered by AID deamination of dC residues (arrows) on the variable or switch regions of the Ig genes, thereby generating dU:dG mismatches on DNA. In SHM the dU:dG mispairing can be replicated over, recognized by UNG, or processed by MMR factors or error-prone polymerases, leading to nucleotide substitutions. In the case of CSR, the dU:dG mismatches introduced by AID in switch regions are recognised by UNG or MMR to produce a substrate for an apyrimidinic endonuclease (ap) that leads to the generation of DSBs. CSR requires switch region DSBs be brought in close proximity, which involves DNA damage response factors including H2AX, ATM, p53-binding protein 1 and Nijmegen syndrome protein 1 and be ligated by non-homologous end joining. CSR results in the deletion of the intervening DNA sequence and the replacement of the C μ constant region by a downstream region (CSR to C γ 1 is shown). In addition, AID-triggered DSBs can also be substrates for pro-oncogenic chromosomal translocations, such as c-myc-IgH translocations found in Burkitt lymphoma. Surveillance pathways mediated by ATM, p53 and ARF provide protective mechanisms against the generation or spreading of AID-promoted chromosome translocations.

allowing the generation of either mutations at A:T pairs by error-prone patch repair or a nicked DNA recombination substrate and (iii) uracil can be removed by uracil-*N*-glycosylase (UNG) to produce an abasic site that enables both transition mutations at G:C pairs and endonuclease cleavage leading to the generation of a DNA double-strand break (DSB) (8). Accordingly, mismatch repair and UNG deficiencies result in shifted mutation patterns towards C:G mutations and transitions at G:C pairs, respectively (8–10). Likewise, CSR is severely decreased in UNG-deficient mice (11) and virtually absent in UNG^{-/-} MutS Homologue 2^{-/-} mice (12). The mechanism that prevents AID-induced deamination from being faithfully repaired by UNG and base excision repair remains unknown, but it might entail a tight coordination of UNG activity and passage of the replication fork, as suggested previously (13). In the case of CSR, generation of DSBs at Ig switch regions is an obligate intermediate leading to the recombination reaction. Switch region DSBs have to be recognized, brought together and ligated with the ensuing deletion of the intervening DNA sequence (Figure 1). CSR ligation is carried out by some components of the non-homologous end joining machinery, including Ku70 and Ku80 subunits of the DNA-dependent protein kinases (14,15). Interestingly, however, neither DNA-dependent protein kinases nor XRCC4 are absolutely required for this reaction (6–19). In addition, a number of protein factors involved in the DNA damage response are important for the resolution of CSR-associated DSBs, including the phosphorylated form of histone variant H2AX (20), the ataxia-telangiectasia-mutated (ATM) protein (21), the p53-binding

protein 1 and Nijmegen syndrome protein 1 component of the Mre11-Rad50-Nbs1 complex (22,23). For a more extensive review on the factors involved in CSR resolution, the reader is referred to references (24,25).

In vitro AID can deaminate cytosines on single-stranded DNA (ssDNA) substrates, whereas double-stranded DNA molecules are refractory to AID activity (26–29). However, transcription renders double-stranded DNA substrates accessible to AID-mediated deamination, both *in vivo* and *in vitro*, implying that transcription can—at least transiently—expose DNA on single-strand configuration (26,29–31). Transcription-associated ssDNA exposure can be achieved in the form of transcription bubbles (30,32), R-loops (33,34) or both. Long before the identification of AID as the initiator activity of the germinal centre reactions, transcriptional activity of variable and switch regions was known to be an absolute requirement for SHM and CSR, respectively. Today, it is accepted that this requirement is mechanistically linked to AID function, since one essential role of Ig transcription during SHM and CSR is to provide AID with access to its bona fide ssDNA substrate.

Regulation of AID function. A number of mechanisms regulate AID function. First, AID expression is mostly restricted to activated B cells. Induction of AID expression is commonly linked to signalling pathways that mediate germinal centre B-cell activation, such as those mimicked by lipopolysaccharide, IL4, transforming growth factor β or CD40 ligation (2,35,36). Accordingly, janus kinase-signal transducer

Table I. Antigen independent AID expression

Cell type	Stimulus	Signalling pathway	Functional effect	Reference
Peripheral blood naive B cells and B cell lines	EBV infection	LMP-1, NF- κ B, BAFF and APRIL	CSR and mutations in BCL-6 and p53 genes	(70,71)
Peripheral blood naive B cells and B cell lines	HCV infection	CD81, CD19, CD21 and TNF α	Mutations in the IgH variable region, BCL-6, p53 and β -catenin genes	(76,77)
Pre-B cells	Ab-MLV infection	NF- κ B	Mutation in the Ig locus	(79)
Immature bone marrow B cells	NA	BCR and TLR	Somatic mutation and CSR in Ig genes	(38,39)
Oocytes, embryonic germ cells and embryonic stem cells	NA	NA	NA	(82)
Spermatocytes	NA	NA	NA	(83)
Epithelial breast cancer cell lines	NA	NA	Somatic mutations in Vh genes and CSR	(88)
Hepatoma cell lines and primary human hepatocellular carcinoma	TGF β , TNF α and HCV infection	NF- κ B	Mutations in p53, c-myc and pim1 genes	(81,85)
Gastric epithelial cells	<i>Helicobacter pylori</i> infection	NF- κ B	Mutations in p53 gene	(87)

APRIL, proliferation-inducing ligand; BAFF, B-cell-activating factor of the tumour necrosis factor family.

and activator of transcription and nuclear factor-kappa B (NF- κ B) pathways seem to be involved in this signalling (35,36). A number of regulatory elements and transcription factors have been described to regulate AID transcription, including the counterbalanced activities of E47 and Id3 and of inhibitory of DNA-binding 2 and Pax5 helix-loop-helix proteins (for a review see ref. 25). Very recently a highly conserved DNA region in the AID gene has been reported essential to allow AID expression *in vivo* as measured in indicator mouse strains (37). In addition, these mouse strains have provided *in vivo* evidence that under physiological conditions, AID is not detected in differentiating bone marrow B cells and that AID expression is induced transiently in germinal centres and is then shut down in fully differentiated plasma cells and memory B cells (37). However, AID expression has been reported in immature B cells from bone marrow (38,39), presumably at levels detectable by PCR but not with indicator mouse strains (37); the significance of this observation requires further examination. Second, AID activity can be regulated by controlling the protein subcellular localization. Indeed, most of AID protein resides in the cytoplasm by virtue of a C-terminal nuclear export signal that restricts its presence in the nucleus (reviewed in ref. 25). Finally, AID activity is enhanced upon phosphorylation by protein kinase A (40–42), possibly by facilitating AID interaction with replication protein A and in turn its access to ssDNA (40,43). Although a lot remains to be understood about these regulatory mechanisms, they are certain to play an important role in minimizing the risk of unwanted DNA damage that AID activity might entail. Of special interest to the topic of this review, we will discuss below some cases of unconventional AID expression and its possible implications in normal and malignant physiology.

Specification of AID targeting. One of the most important issues in understanding the pathogenic consequences of AID function relies on deciphering what is the mechanism that targets AID to the Ig loci and not to other genes throughout the genome (for more extensive reviews on the topic the reader is referred to refs 25,44). AID has preference for so-called mutational hotspots defined as a degenerate trinucleotide motif (A or T, A or G, C) both *in vivo* and *in vitro* (reviewed in ref. 45). Obviously, the low complexity of this motif cannot account for a major specificity of AID activity based on primary DNA sequence alone. Indeed, Ig sequences can be replaced by heterologous sequences that can still support AID activity (reviewed in ref. 46). As mentioned above, transcriptional activation of the Ig loci is an absolute requirement for AID function. Numerous studies have tried to tackle SHM

target specificity on the basis of this feature. Early works showed that removal of *cis* elements at the Ig locus abolishes SHM; however, if transcription is ensured through heterologous regulatory elements, SHM can be achieved (reviewed in refs 44,46). Moreover, AID is known to target genes other than Ig genes: (i) when over-expressed in non-B cell lines, AID can introduce mutations in transgenes as long as they are transcriptionally active, and the rate of SHM correlates with the transcription rate of the target gene (5); (ii) transcriptionally active transgenes are mutated by AID in lymphoma cell lines (47) regardless of their integration site (48) and (iii) SHM-like mutations are found in a number of non-Ig genes of normal B cells, both murine and human, including *BCL6*, *FAS*, *B29* and *MB1* (reviewed in ref. 49), although at frequencies 50- to 100-fold lower than in Ig genes. However, transcription alone is not sufficient to grant AID targeting, as genes known to be transcribed do not accumulate mutations (50). In addition, experiments of IgL promoter replacement in the hypermutating DT40 chicken B-cell line have shown that a strong RNA polymerase II promoter that yields high transcription levels of the Ig locus is nevertheless unable to support SHM (51). This observation could be in line with the finding that an E box in the Ig κ locus enhances SHM without enhancing transcription (52) and with analysis of AID transgenic mice which revealed a common E47-binding motif in the enhancer/promoters of the identified mutated genes (53). In summary, it seems feasible that specification of AID targeting requires a combination of (i) high transcriptional activity and (ii) *cis* elements that could contribute to enhance AID recruitment to particular genes, either directly or through scaffolding by other protein factors. The identity and number of these *cis* elements and any putative protein factors are far from being determined. A more defined picture of the mechanisms regulating AID targeting will probably emerge after comprehensive genome-wide screening of mutated genes and SHM permissive loci.

AID in B-cell lymphomas

A vast majority of human lymphoid malignancies derive from B cells rather than T cells, from which most have their origin in germinal centre or post-germinal centre B lymphocytes. A hallmark of mature B-cell lymphomas is the presence of recurrent chromosome translocations that juxtapose one of the Ig loci and a proto-oncogene, such as c-myc in Burkitt lymphoma, *BCL1* in mantle lymphoma, *BCL6* in diffuse large B cell lymphoma or *BCL2* in follicular lymphomas (reviewed in refs 1,54). As a result of the translocation, expression of the proto-oncogene is deregulated, usually by becoming under the

control of Ig regulatory elements. Deregulated proto-oncogene expression has been proved critical for neoplasia development using *in vivo* models. For instance, transgenic animals expressing c-myc under the control of the IgH enhancer (E μ -myc) develop B-cell lymphomas (55). Translocations that involve the IgH locus are more common than those involving the IgL locus, and break points usually cluster at switch regions or variable regions. Proto-oncogenes are also known to accumulate mutations in some lymphomas (56), (reviewed in ref. 49), but the relevance of these mutations to lymphomagenesis has not been fully determined. Together, these observations suggested from early on that lymphomagenic chromosome translocations could emerge in mature B cells undergoing CSR or SHM, but only after AID identification as the initiator of these reactions, could this hypothesis be tested experimentally.

AID in chromosome translocations. In mice, both IL6 transgenes and pristane injection promote lymphoid hyperplasia and plasmacytosis that are enriched in cells harbouring c-myc/IgH translocations (57). These c-myc/IgH translocations closely resemble those found in human Burkitt lymphoma. In the absence of AID, c-myc/IgH translocations were absent from IL6 transgenic mice (58). A similar result was obtained by Unniraman *et al.* (59), who did not detect c-myc/IgH translocations in AID-deficient mice after pristane-induced plasmacytosis. These results indicated that *in vivo* AID is required for the generation or accumulation of chromosome translocations (for additional discussion on these data the reader is referred to (49)). Additional studies allowed to establish that AID-mediated cytosine deamination can trigger c-myc/IgH translocations *in vitro* in a UNG-dependent fashion, and therefore that the molecular mechanisms that initiate CSR and aberrant chromosome junctions are the same (60).

c-myc/IgH translocations induced upon normal B-cell activation are extremely rare (60,61), which suggests that there might be cellular mechanisms that prevent the generation or spreading of these aberrations and instead drive AID-induced DSBs towards a normal CSR joint. Indeed, ATM seems to be involved in such a surveillance pathway as ATM-deficient B cells accumulate a higher frequency of c-myc/IgH translocations than wild-type cells do. However, the mechanism by which a DSB is recognized by ATM as a translocation intermediate rather than as a CSR substrate remains unknown. Interestingly, frequency of c-myc/IgH translocations increases dramatically in activated B cells that lack p53, suggesting that DSBs that are not resolved into a CSR reaction can be prevented from a translocation event by activation of the p53 pathway through ATM. In addition, ARF-deficient B cells also accumulate a higher frequency of c-myc/IgH translocations, which points to an oncogenic stress response driven by the translocated c-myc that in turn would lead to p53-mediated apoptosis or cellular senescence. Altogether, these results imply that a number of cellular pathways exert protective functions against the appearance of AID-induced chromosome aberrations in B cells. Accordingly, inactivating mutations in p53, ARF and ATM are frequent events in mature B-cell lymphomas (1).

AID in lymphomagenesis. The actual contribution of AID to lymphoma disease development *in vivo* has been more difficult to study. AID transgenic (AIDtg) mice, where AID expression was driven by a ubiquitous chicken β -actin promoter, developed mainly T-cell lymphomas and lung microadenomas (see below) (62). T-cell lymphomas from AIDtg mice accumulated mutations in their T-cell receptor β genes and in the c-myc proto-oncogene, but no translocations were found (62). A similar result was reported by Rucci *et al.* (63) in AIDtg mice under the *lck* T-cell-specific promoter; however, human T-lymphotropic virus promoter failed to promote lymphoma development in spite of similar levels of AID protein being produced. The absence of B-cell lymphomas in mice ubiquitously expressing AID could be due to an earlier onset of T-cell lymphomas. This issue was addressed by the Honjo lab in mice that express AID under a chicken β -actin promoter only when induced through a B-cell-specific

deletion by CD19-Cre expression (64). B-cell lymphomas were not observed in B-cell-specific AIDtg mice up to 20 months of age. Intriguingly, although in these mice the AID transgene produced high levels of AID protein, it failed to fully reconstitute CSR and SHM in AID-deficient mice (64), which suggests that in this model, AID activity might be insufficient to promote the accumulation of lymphomagenic lesions. It is feasible that negative regulatory mechanisms are specifically employed by B cells to restrict the availability of AID function (64). In addition, promoter-specific effects seem critical to determine the outcome of AID transgenic expression (63). In summary, enforced AID expression can promote lymphomagenesis, at least under some experimental conditions, but the reasons for the impaired activity of transgenic AID in some mouse models remain unknown.

What is the lymphomagenic potential of endogenous AID? Mouse models to address this question are scarce. IL6 transgenic mice develop lymphoid hyperplasia regardless of the presence of AID, although AID-deficient mice, where c-myc/IgH translocations are absent, show a delayed onset of the disease (58). This subtle effect can be due to the fact that IL6 over-expression—at least in the absence of an immune response—promotes a polyclonal proliferation and therefore does not completely mimic a lymphoma situation. E μ -myc transgenic mice develop monoclonal pre-B and B-cell lymphomas (55). Kotani *et al.* (65) have recently reported that lymphoma generation in E μ -myc is not dependent on AID. However, AID deficiency induced a dramatic shift in the phenotype of the generated lymphomas, from predominantly mature B cell lymphomas—as measured by the surface expression of Ig—in AID-proficient mice, to pre-B cell lymphomas in AID^{-/-} mice. These results indicate that in the E μ -myc model AID is involved in the generation of mature but not pre-B-cell lymphomas, which is in agreement with AID being expressed at later stages of development. In addition, mutations were found in the *Pim1* proto-oncogene of tumours from AID-proficient mice. These results could be interpreted as AID having a dual role in B-cell development: (i) initiation by chromosomal translocations—provided by the E μ -myc transgene in this model—and (ii) secondary hits by mutagenesis (65).

Novel scenarios for AID expression in B cells. Is deregulated AID expression associated with pro-lymphomagenic situations? AID expression has been reported in germinal centre-derived lymphomas as well as in leukaemias or lymphomas originated from B cells at other stages of differentiation (56,66–69). However, AID expression does not usually correlate with the mutation load of particular proto-oncogenes or with the presence of ongoing SHM in Ig variable regions. This is not surprising, since AID expression may have been functional at earlier stages of the disease and then be shut down. In addition, as pointed out earlier, the presence of AID *per se* may not be warrant of ongoing deamination activity (64).

In addition, induction of AID expression has been reported upon infection with a number of oncogenic viruses. Epstein-Barr virus (EBV) is a gamma-herpesvirus that can promote B-cell transformation and immortalization and is consistently associated with endemic Burkitt lymphoma. Different studies have reported that EBV infection of B cells can induce AID expression (70,71), presumably through one of the EBV gene products, latent membrane protein 1 (LMP-1). LMP-1 cytoplasmic tail is functionally homologous to CD40 and promotes NF- κ B nuclear translocation in a ligand-independent manner (reviewed in ref. 72). He *et al.* (70) showed that LMP-1 expression in lymphoblastoid cells induced AID expression and CSR and promoted the expression of B-cell activation factor of the tumour necrosis factor family and proliferation-inducing ligand, two inducers of T-independent CSR, through NF- κ B activation. More recently, AID and error-prone polymerase η expression has been reported after EBV infection of primary B cells from peripheral blood (71). AID and error-prone polymerase η induction was detected after 10 days of infection and was persistent over long periods of culture, concomitant with the accumulation of mutations in the *BCL6* and p53 proto-oncogenes and in the β -globin gene (71). These results could provide a mechanism explaining the transforming potential of EBV, by means of deregulating AID expression and promoting lymphomagenic

lesions. In apparent contradiction, early Epstein–Barr virus nuclear antigen 2 (EBNA2), another of the nine gene products coded by EBV, has been shown to down-regulate AID expression in lymphoblastoid cell lines (73). This result is striking as EBNA2 in turn induces LMP-1 expression, suggesting that these two gene products might have counteracting effects, that of EBNA2 being dominant over LMP-1 expression. EBV infection can lead to alternative expression programmes; one of them, the so-called latency III state or ‘growth program’, is observed both in lymphoblastoid cells and in infectious mononucleosis and is associated with EBNA2 expression. In contrast, latency I or ‘survival program’ is typical of Burkitt lymphoma cells and is associated with EBNA1 expression and apparently with counter-selection of EBNA2 expression (reviewed in ref. 72). Therefore, EBV infection can result in diverse and complex genetic programmes that are probably intertwined with the nature, differentiation and activation state of the infected cell. Understanding the implications of this complexity for the final outcome of the infection, and most particularly, for AID-associated transforming potential, will demand further investigation.

Hepatitis C virus (HCV) is a causative agent of chronic hepatitis, liver cirrhosis and hepatocellular carcinoma (HCC). In addition to hepatocytes, HCV can also infect B cells both *in vivo* and *in vitro* and occasionally HCV infection is associated with B-cell lymphomas that often harbour chromosome translocations and mutations in the IgH variable region (74,75). Machida *et al.* (76) showed that HCV infection correlated with the accumulation of mutations in various genes, including the IgH variable region, *BCL6* and *p53*, both in cell lines and in peripheral blood B cells from HCV-infected individuals. Moreover, AID and error-prone polymerases were up-regulated upon HCV infection of B cell lines, and AID knockdown by small interfering RNA decreased the mutation load in those genes (76). These effects seem to be exerted by the interaction of the E2 envelope protein of HCV with the CD81 molecule expressed on B cells as part of a signalling receptor complex that includes CD19 and CD21 (77).

Abelson murine leukaemia virus (Ab-MLV) is a highly oncogenic retrovirus that transforms pre-B cells *in vitro* and causes pre-B-cell leukaemia *in vivo* (78). Ab-MLV infection has been shown to promote AID expression in pre-B cells and therefore in a differentiation context well before terminal B-cell maturation and the germinal centre reaction (79). Ab-MLV-induced AID expression results in the accumulation of mutations in the Ig locus in pre-B cells (79). Intriguingly, AID-deficient B cells were found more susceptible to viral transformation by Ab-MLV, suggesting that in this context AID has a safeguarding role against transforming retroviruses, presumably by triggering a genotoxic stress response that restricts the proliferation of the infected cell (79). This is the first evidence that AID could have a protective function against infection independent of antibody diversification but rather as part of an innate defence pathway. In turn, these observations are reminiscent of the known antiviral functions of other cytosine deaminases. Most prominently, apolipoprotein B mRNA-editing enzyme catalytic polypeptide3G, a close relative of AID within the cytosine deaminase family, exerts a protective mechanism against human immunodeficiency virus infection by deamination of the viral DNA generated after reverse transcription (for a review on this topic the reader is referred to ref. 80).

Together, these data indicate that AID-deregulated expression can be tracked to a number of lymphomagenic contexts, not necessarily restricted to the germinal centre, and sometimes underlying viral infection (Table I). Unveiling the relevance of this expression to the actual lymphomagenic process will undoubtedly be the subject of intense research in the near future.

AID in other neoplastic contexts

We have already discussed that AID expression in B cells can lead to the appearance of DNA lesions, most notably chromosome translocations, underlying the aetiology of B-cell lymphomagenesis. Could AID function be linked to the development of non-lymphoid neoplasias?

Two early evidences support the hypothesis that AID is potentially functional in non-B cells; first, heterologous expression of AID in

fibroblasts is sufficient to trigger SHM in transcriptionally active transgenes (5) and second, transgenic mice ubiquitously expressing AID developed T-cell lymphomas and lung microadenomas (62) and, as recently reported, liver tumours in 25% of the cases (81). Therefore, forced AID expression in non-B cells can result in neoplastic development, with particular tissues apparently displaying enhanced susceptibility to malignant transformation.

Interestingly, Morgan *et al.* (82) reported for the first time detection of AID expression in normal tissues other than B cells, in particular in a number of pluripotent tissues, including oocytes, embryonic germ cells and embryonic stem cells, but not in normal somatic tissues such as liver, brain, kidney or heart. More recently, AID expression has also been reported in spermatocytes (83). Although the physiological meaning of AID expression in germ cells is not yet understood, these observations point to a more loosen regulation of AID expression than previously anticipated.

Only very recently has the issue of AID expression in non-lymphoid tumours been addressed. In one instance, AID expression was reported in epithelial breast cancer cell lines (84), although the pathways leading to this aberrant AID expression needs further investigation. In other set of studies, hepatoma cell lines and primary human HCC samples were analysed. HCC is a primary malignancy that very often appears secondary to chronic hepatitis—after hepatitis B or C infection—or to liver cirrhosis. Kou *et al.* examined the presence of AID transcripts in a cohort of samples from normal liver, tissues with chronic liver disease and liver tumours. AID transcripts were detected both in liver tumours and in tissues with chronic liver disease as opposed to normal liver tissues, where only minute AID expression was detectable (85). In addition, the same authors found mutations in the *p53* gene both from tumour and from surrounding non-tumour tissues of HCC patients. Interestingly, the presence of AID transcripts and protein was not attributable to infiltrating B cells alone but was also detected in hepatocytes (85). These results suggest that AID expression could be induced in non-B cells concomitant with pro-inflammatory situations, such as those seemingly characteristic of chronic liver disease leading to hepatocarcinogenesis. Indeed, AID expression can be induced in human hepatocytes after transforming growth factor β (85) or tumour necrosis factor α (81) treatment. In addition, the HCV core protein can also promote AID expression in human hepatocytes. In all cases, AID expression seems dependent on activation of the NF- κ B pathway (81), which is in line with other reports implicating NF- κ B in the transcriptional activation of AID in B cells (36,86). Importantly, these findings establish a correlation between pro-inflammatory situations, for instance those ensuing chronic viral infection, and aberrant AID expression in non-B cells.

Finally, AID expression was analysed in gastric epithelial cells after *Helicobacter pylori* infection (87) that causes chronic gastric inflammation and can eventually result in gastric cancer development. *H. pylori* infection of gastric adenocarcinoma cells promoted the appearance of mutations in the *p53* gene and induced AID expression in an NF- κ B-dependent fashion (87). Interestingly, both effects were achieved only when virulent, *cag* pathogenicity island-positive, *H. pylori* strains were used and *p53* mutation load was decreased after AID knockdown (87). Moreover, AID protein was detected in 78% of *H. pylori*-infected human gastric cancer tissues and *p53* mutations were found in 86% of *H. pylori*-positive gastric cancers and 40% of gastritis specimens (87).

These findings pose a completely new and unexpected picture in which AID deregulation could contribute to a wide spectrum of neoplasias (Table I). However, deciphering the scope of this deregulation and its physiological relevance to malignant transformation will have to await the development of suitable *in vivo* models.

Conclusions and future perspectives

Over the last few years, the identification of AID has allowed a new level of understanding of SHM and CSR, critical reactions for the generation of a highly competent repertoire of antibodies. Concomitant with AID's role in the normal immune response, evidence is

accumulating showing that its activity can also trigger oncogenic lesions, most notably the generation of pro-lymphomagenic chromosome translocations. Moreover, this deleterious effect might not be restricted to B cells alone but instead could strike other cell types upon deregulated expression of AID, for instance in certain pro-inflammatory situations. This novel scenario will certainly raise an intense research effort that will demand the development of suitable *in vivo* models where the contribution of AID to neoplastic development can be assessed. At the same time, understanding the involvement of AID in the generation of DNA lesions will require a minute understanding of its activity and regulation. In this regard, a number of key questions still remain unanswered:

- How AID-mediated cytosine deamination is prevented from faithful repair and is instead driven into mutations or DNA DSBs?
- What are the molecular events leading to AID expression and under what circumstances are AID expression deregulated?
- What is the role of post-transcriptional modifications in the regulation of AID function?
- How is AID target specificity achieved? What is the role of transcription and *cis* regulatory elements in this specificity? When is AID targeting promiscuous and what is the contribution of AID loose specificity to B-cell malignancies?

Funding

Plan National I+D+I (SAF2005-00027) from Spanish Ministerio de Educación y Ciencia.

Acknowledgements

P.P.D. is a Formación de Personal Investigador (FPI) fellow from the Spanish Ministerio de Educación y Ciencia, V.G. de Y. is funded by the Centro Nacional de Investigaciones Oncológicas (CNIO) and A.R.R. is a Ramón y Cajal investigator from the Spanish Ministerio de Educación y Ciencia.

Conflict of Interest Statement: None declared.

References

- Kuppers, R. (2005) Mechanisms of B-cell lymphoma pathogenesis. *Nat. Rev. Cancer*, **5**, 251–262.
- Muramatsu, M. *et al.* (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.*, **274**, 18470–18476.
- Muramatsu, M. *et al.* (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, **102**, 553–563.
- Revy, P. *et al.* (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell*, **102**, 565–575.
- Yoshikawa, K. *et al.* (2002) AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science*, **296**, 2033–2036.
- Okazaki, I.M. *et al.* (2002) The AID enzyme induces class switch recombination in fibroblasts. *Nature*, **416**, 340–345.
- Honjo, T. *et al.* (2005) AID to overcome the limitations of genomic information. *Nat. Immunol.*, **6**, 655–661.
- Petersen-Mahrt, S.K. *et al.* (2002) AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*, **418**, 99–103.
- Rada, C. *et al.* (1998) Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity*, **9**, 135–141.
- Di Noia, J. *et al.* (2002) Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature*, **419**, 43–48.
- Rada, C. *et al.* (2002) AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proc. Natl Acad. Sci. USA*, **99**, 7003–7008.
- Rada, C. *et al.* (2004) Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell*, **16**, 163–171.
- Di Noia, J.M. *et al.* (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.*, **76**, 1–22.
- Casellas, R. *et al.* (1998) Ku80 is required for immunoglobulin isotype switching. *EMBO J.*, **17**, 2404–2411.
- Manis, J.P. *et al.* (1998) Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *J. Exp. Med.*, **187**, 2081–2089.
- Bosma, G.C. *et al.* (2002) DNA-dependent protein kinase activity is not required for immunoglobulin class switching. *J. Exp. Med.*, **196**, 1483–1495.
- Kiefer, K. *et al.* (2007) The catalytic subunit of DNA-protein kinase (DNA-PKcs) is not required for Ig class-switch recombination. *Proc. Natl Acad. Sci. USA*, **104**, 2843–2848.
- Manis, J.P. *et al.* (2002) IgH class switch recombination to IgG1 in DNA-PKcs-deficient B cells. *Immunity*, **16**, 607–617.
- Soulas-Sprauel, P. *et al.* (2007) Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination. *J. Exp. Med.*, **204**, 1717–1727.
- Reina-San-Martin, B. *et al.* (2003) H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *J. Exp. Med.*, **197**, 1767–1778.
- Reina-San-Martin, B. *et al.* (2004) ATM is required for efficient recombination between immunoglobulin switch regions. *J. Exp. Med.*, **200**, 1103–1110.
- Manis, J.P. *et al.* (2004) 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. *Nat. Immunol.*, **5**, 481–487.
- Ward, I.M. *et al.* (2004) 53BP1 is required for class switch recombination. *J. Cell Biol.*, **165**, 459–464.
- Chaudhuri, J. *et al.* (2004) Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.*, **4**, 541–552.
- Ramiro, A. *et al.* (2007) The role of activation-induced deaminase in antibody diversification and chromosome translocations. *Advances in Immunology*. Vol. 94, Academic Press, 75–107.
- Chaudhuri, J. *et al.* (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature*, **422**, 726–730.
- Dickerson, S.K. *et al.* (2003) AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.*, **197**, 1291–1296.
- Bransteitter, R. *et al.* (2003) Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl Acad. Sci. USA*, **100**, 4102–4107.
- Pham, P. *et al.* (2003) Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*, **424**, 103–107.
- Ramiro, A.R. *et al.* (2003) Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat. Immunol.*, **4**, 452–456.
- Shinkura, R. *et al.* (2003) The influence of transcriptional orientation on endogenous switch region function. *Nat. Immunol.*, **4**, 435–441.
- Larijani, M. *et al.* (2007) AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. *Mol. Cell Biol.*, **27**, 20–30.
- Shinkura, R. *et al.* (2004) Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat. Immunol.*, **5**, 707–712.
- Gomez-Gonzalez, B. *et al.* (2007) Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc. Natl Acad. Sci. USA*, **104**, 8409–8414.
- Zhou, C. *et al.* (2003) Human activation-induced cytidine deaminase is induced by IL-4 and negatively regulated by CD45: implication of CD45 as a Janus kinase phosphatase in antibody diversification. *J. Immunol.*, **170**, 1887–1893.
- Dedeoglu, F. *et al.* (2004) Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFκB. *Int. Immunol.*, **16**, 395–404.
- Crouch, E.E. *et al.* (2007) Regulation of AID expression in the immune response. *J. Exp. Med.*, **204**, 1145–1156.
- Mao, C. *et al.* (2004) T cell-independent somatic hypermutation in murine B cells with an immature phenotype. *Immunity*, **20**, 133–144.
- Han, J.-H. *et al.* (2007) Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and toll-like receptors. *Immunity*, **27**, 64–75.
- Basu, U. *et al.* (2005) The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature*, **438**, 508–511.

41. Pasqualucci, L. *et al.* (2006) PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc. Natl Acad. Sci. USA*, **103**, 395–400.
42. McBride, K.M. *et al.* (2006) Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proc. Natl Acad. Sci. USA*, **103**, 8798–8803.
43. Chaudhuri, J. *et al.* (2004) Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature*, **430**, 992–998.
44. Odegard, V.H. *et al.* (2006) Targeting of somatic hypermutation. *Nat. Rev. Immunol.*, **6**, 573–583.
45. Maizels, N. (2005) Immunoglobulin gene diversification. *Annu. Rev. Genet.*, **39**, 23–46.
46. Li, Z. *et al.* (2004) The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev.*, **18**, 1–11.
47. Bachl, J. *et al.* (2001) Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J. Immunol.*, **166**, 5051–5057.
48. Wang, C.L. *et al.* (2004) Genome-wide somatic hypermutation. *Proc. Natl Acad. Sci. USA*, **101**, 7352–7356.
49. de Yebenes, V.G. *et al.* (2006) Activation-induced deaminase: light and dark sides. *Trends Mol. Med.*, **12**, 432–439.
50. Shen, H.M. *et al.* (1998) Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science*, **280**, 1750–1752.
51. Yang, S.Y. *et al.* (2006) Control of gene conversion and somatic hypermutation by immunoglobulin promoter and enhancer sequences. *J. Exp. Med.*, **203**, 2919–2928.
52. Michael, N. *et al.* (2003) The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity*, **19**, 235–242.
53. Kotani, A. *et al.* (2005) A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. *Proc. Natl Acad. Sci. USA*, **102**, 4506–4511.
54. Kuppers, R. *et al.* (2001) Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*, **20**, 5580–5594.
55. Adams, J.M. *et al.* (1985) The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature*, **318**, 533–538.
56. Pasqualucci, L. *et al.* (2001) Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*, **412**, 341–346.
57. Potter, M. (2003) Neoplastic development in plasma cells. *Immunol. Rev.*, **194**, 177–195.
58. Ramiro, A.R. *et al.* (2004) AID is required for c-myc/IgH chromosome translocations *in vivo*. *Cell*, **118**, 431–438.
59. Unniraman, S. *et al.* (2004) Identification of an AID-independent pathway for chromosomal translocations between the IgH switch region and Myc. *Nat. Immunol.*, **5**, 1117–1123.
60. Ramiro, A.R. *et al.* (2006) Role of genomic instability and p53 in AID-induced c-myc-IgH translocations. *Nature*, **440**, 105–109.
61. Roschke, V. *et al.* (1997) Chromosomal translocations deregulating c-myc are associated with normal immune responses. *Oncogene*, **14**, 3011–3016.
62. Okazaki, I.M. *et al.* (2003) Constitutive expression of AID leads to tumorigenesis. *J. Exp. Med.*, **197**, 1173–1181.
63. Rucci, F. *et al.* (2006) Tissue-specific sensitivity to AID expression in transgenic mouse models. *Gene*, **377**, 150–158.
64. Muto, T. *et al.* (2006) Negative regulation of activation-induced cytidine deaminase in B cells. *Proc. Natl Acad. Sci. USA*.
65. Kotani, A. *et al.* (2007) Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. *Proc. Natl Acad. Sci. USA*, **104**, 1616–1620.
66. Greeve, J. *et al.* (2003) Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas. *Blood*, **101**, 3574–3580.
67. Hardianti, M.S. *et al.* (2004) Expression of activation-induced cytidine deaminase (AID) in Burkitt lymphoma cells: rare AID-negative cell lines with the unmutated rearranged VH gene. *Leuk. Lymphoma*, **45**, 155–160.
68. Hardianti, M.S. *et al.* (2004) Activation-induced cytidine deaminase expression in follicular lymphoma: association between AID expression and ongoing mutation in FL. *Leukemia*, **18**, 826–831.
69. Pasqualucci, L. *et al.* (2004) Expression of the AID protein in normal and neoplastic B cells. *Blood*, **104**, 3318–3325.
70. He, B. *et al.* (2003) EBV-encoded latent membrane protein 1 cooperates with BAFF/BLyS and APRIL to induce T cell-independent Ig heavy chain class switching. *J. Immunol.*, **171**, 5215–5224.
71. Epeldegui, M. *et al.* (2007) Infection of human B cells with Epstein-Barr virus results in the expression of somatic hypermutation-inducing molecules and in the accrual of oncogene mutations. *Mol. Immunol.*, **44**, 934–942.
72. Young, L.S. *et al.* (2004) Epstein-Barr virus: 40 years on. *Nat. Rev. Cancer*, **4**, 757–768.
73. Tobollik, S. *et al.* (2006) Epstein-Barr virus nuclear antigen 2 inhibits AID expression during EBV-driven B-cell growth. *Blood*, **108**, 3859–3864.
74. Weng, W.K. *et al.* (2003) Hepatitis C virus (HCV) and lymphomagenesis. *Leuk. Lymphoma*, **44**, 1113–1120.
75. Kitay-Cohen, Y. *et al.* (2000) Bcl-2 rearrangement in patients with chronic hepatitis C associated with essential mixed cryoglobulinemia type II. *Blood*, **96**, 2910–2912.
76. Machida, K. *et al.* (2004) Hepatitis C virus induces a mutator phenotype: enhanced mutations of immunoglobulin and protooncogenes. *Proc. Natl Acad. Sci. USA*, **101**, 4262–4267.
77. Machida, K. *et al.* (2005) Hepatitis C virus E2-CD81 interaction induces hypermutation of the immunoglobulin gene in B cells. *J. Virol.*, **79**, 8079–8089.
78. Shore, S.K. *et al.* (2002) Transforming pathways activated by the v-Abl tyrosine kinase. *Oncogene*, **21**, 8568–8576.
79. Gourzi, P. *et al.* (2006) A role for activation-induced cytidine deaminase in the host response against a transforming retrovirus. *Immunity*, **24**, 779–786.
80. Holmes, R.K. *et al.* (2007) APOBEC-mediated viral restriction: not simply editing? *Trends Biochem. Sci.*, **32**, 118–128.
81. Endo, Y. *et al.* (2007) Expression of activation-induced cytidine deaminase in human hepatocytes via NF- κ B signaling. *Oncogene*.
82. Morgan, H.D. *et al.* (2004) Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.*, **279**, 52353–52360.
83. Schreck, S. *et al.* (2006) Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. *J. Pathol.*, **210**, 26–31.
84. Babbage, G. *et al.* (2006) Immunoglobulin heavy chain locus events and expression of activation-induced cytidine deaminase in epithelial breast cancer cell lines. *Cancer Res.*, **66**, 3996–4000.
85. Kou, T. *et al.* (2007) Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. *Int. J. Cancer*, **120**, 469–476.
86. Gourzi, P. *et al.* (2007) Viral induction of AID is independent of the interferon and the toll-like receptor signaling pathways but requires NF- κ B. *J. Exp. Med.*, **204**, 259–265.
87. Matsumoto, Y. *et al.* (2007) *Helicobacter pylori* infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat. Med.*, **13**, 470–476.
88. Babbage, G. *et al.* (2004) Mantle cell lymphoma with t(11;14) and unmutated or mutated VH genes expresses AID and undergoes isotype switch events. *Blood*, **103**, 2795–2798.

Received August 1, 2007; revised August 27, 2007; accepted August 28, 2007